PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/40, C12N 9/12, 9/16, C12Q 1/18, G01N 33/68 // (C12Q 1/18, C12R 1:725)

A1

(11) International Publication Number: **WO 98/18927**

(43) International Publication Date: 7 May 1998 (07.05.98)

(21) International Application Number: PCT/CA97/00809

(22) International Filing Date: 29 October 1997 (29.10.97)

(30) Priority Data:
60/029,458 30 October 1996 (30.10.96) US

(71) Applicant *(for all designated States except US)*: NATIONAL RESEARCH COUNCIL OF CANADA [CA/CA]; Building M–58, Room EG–12, 1200 Montreal Road, Ottawa, Ontario K1A 0R6 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LEBERER, Ekkehard [DE/CA]; 53 Sweetbriar Drive, Beaconsfield, Québec H9W 5N4 (CA). THOMAS, David, Y. [CA/CA]; 76 Bruck Avenue North, Montreal West, Québec H2X 2E9 (CA).

(74) Agent: COTE, France; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: *CANDIDA ALBICANS* PROTEINS ASSOCIATED WITH VIRULENCE AND HYPHAL FORMATION AND USES THEREOF

(57) Abstract

The present invention relates to *Candida albicans* proteins, such as CaCla4p, Cst20p, CaCdc42p and CaBemlp, associated with virulence and hyphal formation and uses thereof, such as to design screening tests for inhibitors for the treatment of pathogenic fungi infections and/or inflammation conditions. The invention also relates to an *in vitro* screening test for compounds to inhibit the biological activity of at least one protein selected from the group consisting of CaCla4p, Cst20p, CaCdc42p and CaBemlp, which comprises: a) at least one of said proteins; and b) means to monitor the biological activity of said at least one protein; thereby compounds are tested for their inhibiting potential.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

CANDIDA ALBICANS PROTEINS ASSOCIATED WITH VIRULENCE AND HYPHAL FORMATION AND USES THEREOF

BACKGROUND OF THE INVENTION

(a) Field of the Invention

The invention relates to *Candida albicans* proteins, such as CaCla4p, Cst20p, CaCdc42p and CaBemlp, associated with virulence and hyphal formation and uses thereof, such as to design screening tests for inhibitors for the treatment of pathogenic fungi infections and/or inflammation conditions.

(b) Description of Prior Art

Candida albicans is the major fungal pathogen in humans, causing various forms of candidiasis. The incidence of infections is increasing in immunocompromised patients. This fungus is diploid with no sexual cycle and is capable of a morphological transition from a unicellular budding yeast to a filamentous form. Extensive filamentous growth leads to the formation of a mycelium displaying hyphae with branches and lateral buds. In view of the observation that hyphae seem to adhere to and invade host tissues more readily than does the yeast form, the switch from the yeast to the filamentous form probably contributes to the virulence of this organism (for a review see Fidel, P. L. & Sobel, J. D. (1994) *Trends Microbiol.* **2,** 202-205). The molecular mechanisms by which morphological switching is regulated are poorly understood.

Like *C. albicans*, bakers yeast *Saccharomyces cerevisiae* is also a dimorphic organism capable of switching under certain nutritional conditions from a budding yeast to a filamentous form. Under the control of nutritional signals, diploid cells switch to pseudohyphal growth (Gimeno, C. J. et al. (1992) *Cell* **68,** 1077-1090), and haploid cells to invasive growth

(Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8,** 2974-2985).

The similarities between the dimorphic switching of *S. cerevisiae* and *C. albicans* suggest that these morphological pathways may be regulated by similar mechanisms in both organisms. In *S. cerevisiae*, morphological transitions are controlled by signaling components that are also involved in the mating response of haploid cells (Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8,** 2974-2985; Liu, H. et al. (1993) *Science* **262,** 1741-1744). The switch to pseudohyphal growth requires a transcription factor encoded by the *STE12* gene, and a mitogen-activated protein (MAP) kinase cascade including Ste7p (a homolog of MAP kinase kinase or MEK), Ste11p (a MEK kinase homolog) and Ste20p (a MEK kinase kinase) (Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8,** 2974-2985; Liu, H. et al. (1993) *Science* **262,** 1741-1744). The MAP kinases involved in this response are as yet unknown (Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8,** 2974-2985; Liu, H. et al. (1993) *Science* **262,** 1741-1744).

Members of the Ste20p family of serine/threonine protein kinases are thought to be involved in triggering morphogenetic processes in response to external signals in organisms ranging from yeast to mammalian cells. Two of these kinases, Ste20p and Cla4p, are well characterized in *S. cerevisiae* (Leberer, E. et al. (1992) *EMBO J.* **11,** 4815-4824; Cvrckova, F. et al. (1995) *Genes Dev.* **9,** 1817-1830). Ste20p is required for pheromone signal transduction (Leberer, E. et al. (1992) *EMBO J.* **11,** 4815-4824) and for filamentous growth in response to nitrogen starvation (Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8,** 2974-2985; Liu, H. et al. (1993) *Science* **262,** 1741-1744), and shares an essential function with Cla4p during budding (Cvrckova,

F. et al. (1995) *Genes Dev.* **9,** 1817-1830). Ste20p and Cla4p interact with the small G-protein Cdc42p, and this interaction is required for viability of *S. cerevisiae* cells. Ste20p also interacts with the SH3 domain protein Bem1p, and this interaction plays a role in morphogenetic processes (Leeuw, T. et al. (1995) *Science* **270,** 1210-1213).

Here we show that Cst20p, a *C. albicans* homolog of the Ste20p protein kinase, is required for hyphal growth of *C. albicans* under certain *in vitro* conditions. We also show in a mouse model for systemic candidiasis that Cst20p plays a role in virulence, as judged from significantly prolonged survival of mice infected with *CST20* deleted cells. Our results suggest that Cst20p acts in a regulatory pathway which is involved in hyphal growth of *C. albicans*.

We also demonstrate that CaCla4p, a *C. albicans* homolog of the Cla4p protein kinase, is required for hyphal formation *in vitro* in response to serum, and *in vivo* in a mouse model for systemic candidiasis. We also show that CaCla4p is required for efficient colonization of kidneys with *C. albicans* cells after infection of mice and essential for virulence in the mouse model.

SUMMARY OF THE INVENTION

One aim of the present invention is to provide *Candida albicans* proteins, such as CaCla4p, Cst20p, CaCdc42p and CaBem1p, and their uses thereof.

One aim of the present invention is to provide the nucleotide and amino acid sequences of CaCla4p, Cst20p, CaCdc42p and CaBem1p.

Another aim of the present invention is to provide screening tests for inhibitors of CaCla4p, Cst20p, CaCdc42p and CaBem1p or of their interactions.

The term "fungi" when used herein is intended to mean any fungi, pathogenic or not, which show hyphal induction using kinases, such as *C. albicans*, *Saccharomyces cerevisiae*, *Aspergillus*, *Ustilago maydis*, and all the species of the fungal genera *Aspergillus*, *Blastomyces*, *Candida*, *Cladosporium*, *Coccidioides*, *Cryptococcus*, *Epidermophyton*, *Exophilia*, *Fonsecaea*, *Histoplasma*, *Madurella*, *Malassezia*, *Microsporum*, *Paracoccidioides*, *Penicillium*, *Phaeoannellomyces*, *Phialophora*, *Scedosporium*, *Sporothrix*, *Torulopsis*, *Trichophyton*, *Trichosporon*, *Ustilago*, *Wangiella*, *Xylohypha*, among others.

In accordance with the present invention there is provided an *in vitro* screening test for compounds to inhibit the biological activity of at least one protein selected from the group consisting of CaCla4p, Cst20p, Cdc42p and Bemlp, which comprises:

a) at least one of the proteins; and

b) means to monitor the biological activity of at least one protein;

thereby compounds are tested for their inhibiting potential.

In accordance with another embodiment of the present invention, the inhibition of the interactions between CaCla4p and CaCdc42p is determined.

In accordance with another embodiment of the present invention, the inhibition of the interactions between Cst20p and CaCdc42p is determined.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A to 1D illustrate photomicrographs which show that *C. albicans CST20* gene complements defects in pseudohyphal growth of *ste20/ste20 S. cerevisiae* diploid cells.

Figs. 2A to 2C show the morphology of *S. cerevisiae MAT*α cells (strain YEL306-1A) deleted for *STE20* and *CLA4*, and transformed with plasmids expressing *CLA4*

(Fig. 2A), *STE20* (Fig. 2B) and *C. albicans CST20* (Fig. 2C).

Figs. 3A to 3C show the nucleotide (SEQ ID NO:5) and predicted amino acid sequences of *CST20* (SEQ ID NO:6).

Fig. 4A is the deletion of *CST20* in *C. albicans*.

Fig. 4B is the Southern blot analysis with a *CST20* fragment from *Eco*RI to *Xba*I as a probe.

Figs. 5A to 5J show colonies of *C. albicans* cells grown for 5 days at 37°C on solid "Spider" medium containing mannitol. Wild type strain SC5314 (**A**), *ura3/ura3 cst20Δ/cst20Δ::URA3* strain CDH22 (**B**), *ura3/ura3 cst20Δ/cst20Δ::CST20::URA3* strain CDH36 (obtained by reintegration of *CST20* into strain CDH25 by homologous recombination using linearized plasmid pDH190) (**C**), *ura3/ura3 cst20Δ/cst20Δ* strain CDH25 transformed with plasmids pYPB1-ADHpt (**D**) and pYPB1-ADHpt-HST7 (**E**), *ura3/ura3 hst7Δ/hst7Δ* strain CDH12 transformed with plasmids pVEC (**F**), pVEC-HST7 (**G**), pYPB1-ADHpt (**H**), and pYPB1-ADHpt-HST7 (**I**), and *ura3/ura3 cph1/cph1* strain CDH72 [*ura3/ura3* derivative of strain JK19] transformed with pYPB1-ADHpt-HST7 (**J**). Photomicrographs of representative colonies were taken with a 2x lens (bar=2mm).

Figs. 6A to 6C illustrate virulence assays. Survival curves of mice (n=10 for each *C. albicans* strain at each inoculation dose) infected with 1×10^6 (**A**) and 1×10^5 (**B**) cells of *C. albicans* strains SC5314 (wild type), CAI4 (*ura3/ura3*), CDH22 (*ura3/ura3 cst20Δ/cst20Δ ::URA3*) (**C**) Staining of mouse kidney sections with periodic acid Schiff's stain 48 hours after infection with *cst20Δ/cst20Δ::URA3* mutant strain CDH22 (**a**). Some hyphal cells are indicated with arrows (bar=0.1 mm).

Figs. 7A to 7B illustrate the nucleotide (SEQ ID NO:7) and predicted amino acid (SEQ ID NO:8) sequences of *CaCLA4*.

Fig. 8A illustrates the deletion of *CaCLA4* in *C. albicans*.

Fig. 8B illustrates the Southern blot analysis with the *CaCLA4* fragment from *Pst*I to *Xba*I as a probe.

Fig. 8C illustrates the Northern blot analysis with the *CaCLA4* fragment as a probe. PCR with the divergent oligodeoxynucleotides OEL109 and OEL110 was used to delete the coding sequence of *CaCLA4*. A *hisG-URA3-hisG* cassette was then inserted, and homologous recombination was used in a two-step procedure to replace both *CaCLA4* alleles.

Fig. 9 illustrates virulence assays. Survival curves of mice (n=15 for each *C. albicans* strain) infected with 1 x 10^6 cells of *C. albicans* strains SC5314 (wild-type), CDH77 (*CaCLA4/cacla4Δ*), CLJ1 (*cacla4Δ/cacla4Δ*) and CLJ5 (*CaCla4Δ/cacla4Δ*) transformed with the control plasmid pVEC and plasmid pVEC-*CaCLA4* carrying the *CaCLA4* gene.

Fig. 10 illustrates the staining of mouse kidney sections with periodic acid Schiff's stain 48 h after infection with *C. albicans* strains SC5314 and CLJ1.

Fig. 11 illustrates the nucleotide (SEQ ID NO:9) and predicted amino acid (SEQ ID NO:10) sequences of CaCdc42p.

Figs. 12A to 12B illustrate the nucleotide (SEQ ID NO:11) and predicted amino acid (SEQ ID NO:12) sequences of CaBem1p.

DETAILED DESCRIPTION OF THE INVENTION

The *CST20* gene of *Candida albicans* was cloned by functional complementation of a deletion of the *STE20* gene in *Saccharomyces cerevisiae*. *CST20* encodes a homolog of the Ste20p/p65^{PAK} family of protein kinases.

Colonies of *C. albicans* cells deleted for *CST20* revealed defects in the lateral formation of mycelia on synthetic solid "Spider" media. However, hyphal development was not impaired in some other media. Cells deleted for *CST20* were less virulent in a mouse model for systemic candidiasis. Our results suggest that more than one signaling pathway can trigger hyphal development in *C. albicans*, one of which has a protein kinase cascade that is analogous to the mating response pathway in *S. cerevisiae* and might have become adapted to the control of mycelial formation in asexual *C. albicans*.

The *CaCLA4* gene of *C. albicans* was cloned by functional complementation of the growth defect of *S. cerevisiae* cells deleted for the *STE20* and *CLA4* genes. *CaCLA4* encodes a homolog of the Ste20p family of serine/threonine protein kinases with pleckstrin homology and Cdc42p binding domains in the amino-terminal non-catalytic region. Deletion of both alleles of *CaCLA4* in *C. albicans* caused defects in hyphal formation *in vitro* in synthetic liquid and solid media, and *in vivo* in a mouse model for systemic candidiasis. The deletions reduced the invasion of *C. albicans* cells into kidneys after infection into mice and completely suppressed virulence in the mouse model. Thus, hyphal formation of *C. albicans* mediated by the CaCla4p protein kinase may contribute to the pathogenicity of this dimorphic fungus.

The *CaBEM1* and *CaCDC42* genes of *C. albicans* were cloned by functional complementation of the growth defect of *S. cerevisiae* cells deleted for the *BEM1* and *CDC42* genes, respectively. *CaBEM1* encodes an SH3 domain protein with homology to Bemlp, and *CaCDC42* encodes a small G-protein with homology to members of the Rho-family of G-proteins.

MATERIALS AND METHODS

Yeast manipulations

The yeast form of *C. albicans* was cultured at 30°C in YPD medium. Hyphal growth was induced at 37°C on solid "Spider" media (Liu, H. et al. (1994) *Science* **266,** 1723-1726) containing 1% (w/v) nutrient broth, 0.2% (w/v) K_2HPO_4, 2% (w/v) agar and 1% (w/v) of the indicated sugars (pH 7.2 after autoclaving). Cells were grown in liquid "Spider" media at 30°C to stationary phase, and then incubated for 5 days at 37°C on solid "Spider" media at a density of about 200 cells per 80 mm plates. All media were supplemented with uridine (25 µg/ml) for the growth of Ura$^-$ strains. Germ tube formation was induced at 37°C in either 10% fetal bovine serum (GIBCO/BRL) on liquid "Spider" media containing the indicated sugars at an inoculation density of 10^7 cells per ml.

Yeast manipulations were performed according to standard procedures.

Isolation of *CST20*

The *CST20* gene was isolated from a genomic *C. albicans* library constructed in plasmid YEp352 from genomic DNA of the clinical isolate WO1 (Boone, C. et al. (1991) *J. Bacteriol.* **173,** 6859-6864). A plasmid carrying an amino-terminally truncated version of *CST20* missing the first 918 nucleotides of coding sequence was isolated by screening for suppressors of defects in basal *FUS1::HIS3* expression and mating in *S. cerevisiae* strain YEL64 which was disrupted in *STE20*. A fragment from nucleotides 958 to 1,252 of *CST20* was amplified by the polymerase chain reaction (PCR) and used as a probe to isolate a full length clone by colony hybridization to the *C. albicans* genomic library transformed into *E. coli* strain MC1061. Both DNA strands were sequenced by

the dideoxy chain termination method. The full length clone was subcloned between the *Sac*I and *Hind*III sites of the *S. cerevisiae* centromere plasmid pRS316 to yield plasmid pRL53.

Isolation of *CaCLA4*

The *S. cerevisiae MAT*α strain YEL257-1A-2 deleted for *STE20* and *CLA4* and carrying plasmid pDH129 with *CLA4* under control of the *GAL1* promoter was transformed with the genomic *C. albicans* library constructed in the *S. cerevisiae* vector YEp352 carrying *URA3* as selectable marker (Boone, C. et al. (1991) *J. Bacteriol.* **173,** 6859-6864). Transformants were grown on selective medium in 4% galactose and then replica-plated to selective medium containing 2% glucose to select for plasmids that were able to support growth in the absence of Cla4p and Ste20p. By screening 1,600 transformants, we isolated plasmid YEp352-CaCLA4 carrying an insert of 5.6 kb with an open reading frame of 2,913 bp capable of encoding a homolog of Cla4p. Subcloning indicated that this open reading frame was responsible for complementation. Both DNA strands were sequenced by the dideoxy chain termination method.

Molecular cloning of *CaCDC42*

The *S. cerevisiae MAT*α strain DJTD2-16A carrying the *cdc42-1*^{ts} mutation was transformed with the genomic *C. albicans* library constructed in the *S. cerevisiae* vector YEp352 carrying *URA3* as selectable marker (Boone, C. et al. (1991) *J. Bacteriol.* **173,** 6859-6864). Transformants were grown on selective medium at room temperature. Colonies were then replica-plated to selective medium and grown at 34°C. By screening 2,000 transformants, we isolated plasmid YEp352-CaCDC42 carrying an open reading frame of 573 bp capable of encoding a homolog of Cdc42p. Both DNA strands were sequenced by the dideoxy chain termination method. Sub-

cloning of various restriction endonuclease fragments indicated that the open reading frame was responsible for complementation of the temperature-sensitive growth defect caused by the *cdc42-1^{ts}* mutation.

Molecular cloning of *CaBEM1*

The *S. cerevisiae MATα* strain YEL220-1A deleted for *BEM1* and carrying plasmid pGAL-BEM1 with *BEM1* under control of the *GAL1* promoter was transformed with the genomic *C. albicans* library constructed in the *S. cerevisiae* vector YEp352 carrying *URA3* as selectable marker (Boone, C. et al. (1991) *J. Bacteriol.* **173**, 6859-6864). Transformants were grown on selective medium in 4% galactose and then replica-plated to selective medium containing 2% glucose to select for plasmids that were capable of supporting growth of Bemlp-depleted cells. We isolated plasmid YEp352-CaBEM1 carrying an open reading frame of 1,905 bp fulfilling this criterion and capable of encoding a homolog of Bemlp. Both DNA strands were sequenced by the dideoxy chain termination method, and subcloning of various restriction endonuclease fragments indicated that this open reading frame was responsible for complementation.

Construction of *C. albicans* strains and plasmids

To construct a *CST20* null mutant, an *Eco*RI to *Sac*I fragment from nucleotide positions 989 to 4,134 of *CST20* was subcloned into the Bluescript KS(+) vector (Stratagene) to yield plasmid pDH119. A plasmid that contained *CST20*-flanking sequences from nucleotides 989 to 1,674, and 3,423 to 4,134 joined with *Bam*HI sites, was then created by PCR using the divergent oligodeoxynucleotide primers ODH68 (5'-CGGGATCCAGACCAACCACTCGAACTACT-3' (SEQ ID NO:1) and ODH69 (5'-CGGGATCCGAAGGTGAACCACCATATTTG-3' (SEQ ID NO:2); newly introduced *Bam*HI sites are underlined) and plasmid pDH119 as a template. The amplified DNA was

cleaved with *Bam*HI and ligated with a 4 kb *Bam*HI to *Bgl*II fragment of a *hisG-URA3-hisG* cassette derived from plasmid pCUB-6 (Fonzi, W. A. & Irwin, M. Y. (1993) *Genetics* **134,** 717-728) to yield plasmid pDH183. This plasmid was linearized with *Xho*I and *Sac*I and transformed into the Ura$^-$ *C. albicans* strain CAI4 (Fonzi, W. A. & Irwin, M. Y. (1993) *Genetics* **134,** 717-728) to partially replace the coding region of one of the chromosomal *CST20* alleles with the *hisG-URA3-hisG* cassette by homologous recombination. Ura$^+$ transformants were selected on Ura$^-$ medium, and integration of the cassette into the *CST20* locus was verified by Southern blot analysis. Spontaneous Ura$^-$ derivatives of two of the heterozygous disruptants were selected on medium containing 5-fluoroorotic acid. These clones were screened by Southern blot hybridization to identify those which had lost the *URA3* gene by intrachromosomal recombination mediated by the hisG repeats. This procedure was then repeated to delete the remaining functional allele of *CST20*.

A similar procedure was employed to delete the *CaCST20* gene. A 4.6 kb *Xba*I fragment of YEp352-*CaCLA4* was subcloned into the pBluescript KS(+) vector (Stratagene) to yield plasmid pDH205. A plasmid that contained *CaCLA4* flanking sequences joined with *Bgl*II sites was then created by PCR using the divergent oligodeoxynucleotide primers OEL109 (5'-GAAGATCTTGTAATCAATGTTCCCGTGGA-3' (SEQ ID NO:3) and OEL110 (5'-GAAGATCTCATCGTGATATTAAATCCGAT-3' (SEQ ID NO:4); newly introduced *Bgl*II sites are underlined) and plasmid pDH205 as template. The amplified DNA was cleaved with *Bgl*II and ligated with a 4 kb *Bam*HI-*Bgl*II fragment of a *hisG-URA3-hisG* cassette derived from plasmid pCUB-6 (Fonzi, W. A. & Irwin, M. Y. (1993) *Genetics* **134,** 717-728) to yield plasmid pDH210. This

plasmid was linearized with *Pst*I and *Sac*I and transformed into the Ura$^-$ *C. albicans* strain CAI4 (Fonzi, W. A. & Irwin, M. Y. (1993) *Genetics* **134,** 717-728) to replace the coding region of one of the chromosomal *CaCLA4* alleles with the *hisG-URA3-hisG* cassette by homologous recombination. Ura$^+$ transformants were selected on Ura$^-$ medium, and integration of the cassette into the *CaCLA4* locus was verified by Southern blot analysis. Spontaneous Ura$^-$ derivatives were then selected on medium containing 5-fluoroorotic acid. These clones were screened by Southern blot hybridization to identify those which had lost the *URA3* gene by intrachromosomal recombination mediated by the *hisG* repeats. This procedure was then repeated to delete the remaining functional allele of *CaCLA4*.

To reintegrate *CST20* into the genome of mutant strains, the *C. albicans* integration plasmid pDH190 was constructed by subcloning a *Kpn*I to *Pst*I fragment of *CST20* into pBS-c*URA3* (pBluescript KS(+) into which the *C. albicans URA3* gene was cloned between the *Not*I and *Xba*I sites of the polylinker). The integration plasmid was then linearized with *Nsi*I and transformed into *C. albicans* to target integration into the *Nsi*I site of the *CST20Δ*::hisG fusion gene. Integrations were selected on Ura$^-$ medium and confirmed by Southern blot analysis.

The *C. albicans CST20* expression plasmid pDH188 was constructed by subcloning a *Sac*I to *Pst*I fragment of *CST20* into plasmid pVEC carrying a *C. albicans* autonomously replicating sequence and *URA3* as selectable marker. The *C. albicans* plasmid pVEC-*CaCLA4* was constructed by subcloning the *Kpn*I to *Sac*I insert of YEp 352-*CaCLA4* into plasmid pVEC.

Northern blot analyses

Northern blots of total and poly $(A)^+$ RNA from *C. albicans* cells were performed as described (Leberer, E. et al. (1992) *EMBO J.* **11**, 4815-4824). Signals were quantified by 2-D radioimaging.

Animal experiments

Eight week-old, male CFW-1 mice (Halan-Winkelmann, Paderborn, Germany) were inoculated with 1×10^5 or 1×10^6 cells by intravenous injection. Survival curves were calculated according to the Kaplan-Meier method using the PRISM™ program (GraphPad Software Inc., San Diego) and compared using the log-rank test. A P value <0.05 was considered significant.

To quantify colony-forming *C. albicans* units in kidneys, mice were sacrificed by cervical dislocation 48 hours after injection and kidneys were homogenized in 5 ml phosphate buffered saline, serially diluted and plated on YNG medium (0.67% yeast nitrogen base, 1% glucose, pH 7.0). Histological examination of kidney sections was done with periodic acid Schiff's stain.

RESULTS

Isolation and characterization of *CST20*

A *C. albicans* homolog of the *S. cerevisiae STE20* gene was cloned by functional complementation of the pheromone signaling defect of *S. cerevisiae* cells that were deleted for the *STE20* gene. The mating defect of the *STE20* deleted *S. cerevisiae* strain YEL20 was fully complemented by introduction of the centromeric plasmid pRL53 carrying full length *CST20* (mating efficiency was 81±9% in cells expressing *CST20*, compared with 85±8% in cells expressing *STE20*; n=3). Similarly, defects in growth arrest and morphological changes in response to pheromone were completely cured by transformation with the *CST20* plasmid.

As shown in Fig. 1, nitrogen deficiency-induced pseudohyphae formation, which is blocked by disruption of *STE20* in diploid cells (Liu, H., Styles, C. & Fink, G. R. (1993) *Science* **262,** 1741-1744), was restored by introduction of the *CST20* plasmid. Colonies of the diploid *STE20* wild type strain L5266 (4) (Fig. 1A) and the isogenic *ste20/ste20* strain HLY492 (4) transformed with either the control plasmid pRS316 (Fig. 1B), the *CST20* plasmid pRL53 (Fig. 1C), or the *STE20* plasmid pSTE20-5 (9) (Fig. 1D) were grown on nitrogen starvation medium (2) for 5 days at 30°C. Photomicrographs were taken with a 4x objective (bar=1mm).

As illustrated in Fig. 2, the cytokinesis defect caused by deletion of *CLA4*, encoding an *S. cerevisiae* isoform of Ste20p (Cvrckova, F. et al. (1995) *Genes Dev.* **9,** 1817-1830), was not complemented by *CST20* (Fig. 2). However, the lethality caused by deletion of both *STE20* and *CLA4* (Cvrckova, F. et al. (1995) *Genes Dev.* **9,** 1817-1830), could be rescued by *CST20* (Fig. 2). The diploid strain YEL306 heterozygous for *ste20Δ ::TRP1/STE20 cla4Δ::LEU2/CLA4* was transformed with plasmid pRS316 carrying either no insert, *CLA4* (pRL21), *CST20* (pRL53) or *STE20* (pSTE20-5), and then sporulated and dissected. No viable haploid *ste20Δ cla4Δ* spores were obtained from transformants with the plasmid without insert, but were obtained from transformants with plasmids carrying *CLA4* (Fig. 2A), *STE20* (Fig. 2B) or *CST20* (Fig. 2C).

Cells were grown to mid-exponential phase in YPD medium at 30°C. No viable *ste20Δ cla4Δ* segregants were obtained in medium containing 5-fluoro-orotic acid suggesting that the plasmids were essential for viability. Neither *STE20* nor *CST20* were able to suppress the morphological defect of *cla4Δ* cells. Photomicrographs

were taken by phase contrast with a 40x objective (bar=30 μm).

The open reading frame of *CST20* is capable of encoding a protein of 1,229 amino acids with a predicted molecular weight of 133 kDa and a domain structure characteristic of the Ste20p/p65^{PAK} family of protein kinases (Fig. 3). Numerals at the left margin indicate nucleotide and amino acid positions (Fig. 3). Nucleotide 1 corresponds to the first nucleotide of the initiation codon and amino acid 1 to the first residue of the deduced protein. The putative p21 binding domain has been shadowed, and the kinase domain has been boxed.

The catalytic domain present in the carboxyl terminal half of the protein has sequence identities of 76 and 56%, respectively, with *S. cerevisiae* Ste20p (Leberer, E. et al. (1992) *EMBO J.* **11**, 4815-4824) and Cla4p (Cvrckova, F. et al. (1995) *Genes Dev.* **9**, 1817-1830). The amino terminal, non-catalytic region contains a sequence from amino acid residues 473 to 531 with 68% identity to the p21 binding domain of Ste20p that has been shown to bind the small GTPase Cdc42p. This region contains the sequence motif ISxPxxxxHxxH thought to be important for the interaction of the p21 binding domain with the GTP-bound forms of Cdc42Hs and Racl (Cvrckova, F. et al. (1995) *Genes Dev.* **9**, 1817-1830). The remaining non-catalytic sequences are less conserved. Unique sequences not present in Ste20p and the other members of the family are found at the amino terminus and between the p21 binding and catalytic domains.

A *CST20* transcript of 4.9 kb in size was detected in Northern blots. This transcript was present at similar levels in yeast cells grown in YPD at

room temperature and germ tubes induced by a temperature shift to 37°C.

Isolation and characterization of *CaCLA4*

A *C. albicans* homolog of the *S. cerevisiae CLA4* gene was cloned by functional complementation of the growth defect of *S. cerevisiae* cells that were deleted for the *STE20* and *CLA4* genes.

The open reading frame of the *CaCLA4* gene is capable of encoding a protein of 971 amino acids with a predicted molecular weight of 107 kDa and a domain structure characteristic of the Ste20p family of protein kinases (Fig. 7). The catalytic domain present in the carboxyl terminal half of the protein has sequence identities of 74, 63 and 64%, respectively, with *S. cerevisiae* Cla4p, *S. cerevisiae* Ste20p and an uncharacterized open reading frame present in the *S. cerevisiae* genome, 65% with the *C. albicans* Ste20p homolog Cst20p, and 61% with rat $p65^{PAK}$ (Fig. 7). The amino terminal, noncatalytic region contains a sequence from amino acid residues 69 to 180 with similarity to pleckstrin homology (PH) domains and a sequence from amino acid residues 229 to 292 with 63% identity to the Cdc42p binding domain of *S. cerevisiae* Cla4p that has been shown to bind the small GTPase Cdc42p (Cvrckova, F. et al. (1995) *Genes Dev.* **9**, 1817-1830). The remaining noncatalytic sequences are less conserved.

Chromosomal deletion of *CST20*

Homologous recombination was used in a multistep procedure to partially delete *CST20* in a URA^- *C. albicans* strain (Fig. 4A). PCR with the divergent oligodeoxynucleotides ODH68 and ODH69 was used to partially delete the coding sequence of *CST20*. A *hisG-URA3-hisG* cassette was then inserted. The deletion was confirmed by Southern blot analyses (Fig. 4B). The genomic DNA samples digested with *Xho*I were from following strains:

Lane #1, CAI4 (*ura3/ura3 CST20/CST20*); lane 2, CDH15 (*ura3/ura3 CST20/cst20Δ::hisG-URA3-hisG*); lane 3, CDH18 (*ura3/ura3 CST20/cst20Δ::hisG*); lane 4, CDH22 (*ura3/ura3 cst20Δ::hisG-URA3-hisG/cst20Δ::hisG*); lane 5, CDH25 (*ura3/ura3 cst20Δ::hisG/cst20Δ::hisG*). Northern blots showed that the *CST20* transcript was absent in the corresponding homozygous deletion strains.

The lateral outgrowth of hyphae from colonies grown on solid "Spider" media containing mannitol or sorbitol was completely blocked by deletion of *CST20* (Fig. 5B).

Mycelial formation was drastically reduced when the media contained galactose, mannose or raffinose. The mutant strains regained the ability to form hyphae when wild type *CST20* was reintroduced by transformation with the *CST20* expression plasmid pDH188 or reintegrated into the genome by targeted homologous recombination (Fig. 5C). The *CST20* transcript was detected in these strains by Northern blot analysis.

Mutant strains formed hyphae when colonies were grown on "Spider" media containing either glucose or N-acetyl glucosamine. Normal hyphae formation was also observed on rice agar and on agar containing Lee's medium or 10% serum. The frequency of germ-tube formation in either liquid Lee's medium, 10% serum or liquid "Spider" media containing any of the sugars tested above, were also normal. These results indicate that Cst20p is not required for hyphae formation under all conditions but are involved in the lateral formation of mycelia on some solid surfaces.

Chromosomal deletion of *CaCLA4*

Homologous recombination was used in a multistep procedure to delete both alleles of *CaCLA4* in *C. albicans* (Fig. 8A). Fig. 8A shows the restriction endonuclease map of *CaCLA4*. The coding sequence is indicated

by the arrow. PCR with the divergent oligodeoxynucleotides OEL109 and OEL110 was used to delete the coding sequence of *CaCLA4*. A *hisG-URA3-hisG* cassette was then inserted and a two-step procedure was used to delete both alleles of *CaCLA4* by homologous recombination. The endonuclease restriction sites are as follows: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sac*I; X, *Xba*I. The deletions were confirmed by Southern blot analyses (Fig. 8B). Southern blot analysis with a 1.1 kb *CaCLA4* fragment from *Pst*I-*Xba*I as a probe. The genomic DNA samples digested with *Eco*RI were from following strains: Lanes: 1, CAI4 (*ura3/ura3 CaCLA4/CaCLA4*); 2, CDH77 (*ura3/ura3 CaCLA4/cacla4Δ ::hisG-URA3-hisG*); 3, CDH88 (*ura3/ura3 CaCLA4/cacla4Δ ::hisG*); 4, CLJ1 (*ura3/ura3 cacla4Δ::hisG-URA3-hisG/cacla4Δ::hisG*); and 5, CLJ5 (*ura3/ura3 cacla4Δ ::hisG/cacla4Δ::hisG*). Northern blots showed that the *CaCLA4* transcript with a size of 4.1 kb was reduced to about 40% in heterozygous *CaCLA4/cacla4Δ* cells and was absent in homozygous *cacla4Δ/cacla4Δ* deletion cells (Fig. 8C). The transcript was present at about wild-type levels when the *CaCLA4* gene was retransformed into the homozygous deletion cells by using an autonomously replicating plasmid carrying the *CaCLA4* gene (Fig. 8C). Northern blot analysis of poly(A)$^+$ RNA isolated from following strains grown in the yeast form in YPD at 30°C: Lanes: 1, SC5314 (wild-type); 2, CDH88; 3, CLJ5 transformed with pVEC; 4, CLJ5 transformed with pVEC-CaCLA4. The blot was probed with fragments specific for *CaCLA4* (upper panel) or *CaACT1* (lower panel) and quantified by radioimaging. Numbers at the bottom of the figure depict the relative amounts of *CaCLA4* transcript in relation to the amounts of *CaACT1* transcript (mean values of two independent experiments).

We found that viability of *C. albicans* cells was not affected by deleting either one or both alleles of *CaCLA4*. Mutant cells showed the same growth behavior as wild-type cells, independently whether the cells were grown under conditions favoring either the yeast or filamentous forms. However, deletion of both *CaCLA4* alleles generated defects in cellular morphology producing a heterogeneous population of aberrantly shaped cells that were frequently multibudded and multinucleated. This phenotype indicates a defect in cytokinesis resembling the phenotype of *S. cerevisiae* cells deleted for *CLA4* (Cvrckova, F. et al. (1995) *Genes Dev.* **9**, 1817-1830).

Deletion of both *CaCLA4* alleles caused defects in hyphal formation in all media and under all conditions that we investigated. When morphological switching was induced in liquid media by either serum, N-acetyl glucosamine, proline, pH increase, temperature shift, or Lee's medium, wild-type cells and cells deleted for only one or both alleles of *CaCLA4* produced germ tubes after about 30 minutes. In wild-type cells and cells deleted for only one allele of *CaCLA4*, these germ tubes elongated and grew into long hyphae after prolonged incubation. Cells deleted for both alleles of *CaCLA4* failed to produce hyphae, however. Instead, these cells produced multiple short protrusions giving rise to an aberrant morphology.

On solid media containing either serum, rice agar or mannitol, the normal formation of mycelia was completely suppressed by deletion of both *CaCLA4* alleles. This phenotype was reversed by introducing the *CaCLA4* gene on a plasmid, and deletion of only one allele had no effect.

Virulence studies

To determine the role of Cst20p for virulence, mice were injected intravenously with wild type and mutant strains and monitored for survival and for fungal invasion into kidneys. We found that the Ura$^-$ strain CAI4 was not pathogenic (Figs. 6A and B). However, infection with Ura$^+$ wild type cells resulted in rapid mortality with a rate that was dependent on the dose of injected cells (1 x 10^6 cells in Fig. 6A, and 1 x 10^5 cells in Fig. 6B). Survival was significantly prolonged, however, in mice infected with Ura$^+$ cells deleted for both alleles of *CST20* (*cst20Δ/cst20Δ::URA3*). This effect, which was reproducible and statistically significant, was observed at high (Fig. 6A) or low (Fig. 6B) doses of infection (with P values of 0.027 and 0.001, respectively) and correlated with colony-forming units per kidney (1.5 x 10^6 for wild type cells and 7 x 10^5 for *cst20Δ/cst20Δ::URA3* mutant cells) after 48 hours of infection with 1 x 10^6 cells. These effects on virulence could be reversed by reintroducing *CST20* into the strain deleted for both *CST20* alleles, and were not observed in Ura$^+$ cells deleted for only one *CST20* allele. A histological examination revealed that cells deleted for both alleles of *CST20*, were able to form hyphae in infected kidneys (Fig. 6C).

To investigate whether CaCla4p is required for virulence, mice were injected intravenously with wild-type and mutant *C. albicans* strains and monitored for survival and for fungal invasion into kidneys. Infections with *CaCLA4* wild-type cells (strain SC5314) resulted in rapid mortality (Fig. 9). No difference in the mortality rate was observed after infection with cells deleted for only one allele of *CaCLA4* (strain CDH77). All mice survived, however, after infection with cells deleted for both alleles of *CaCLA4* (strain

CLJ1 and CLJ5pVEC1). This effect correlated with a reduction in the amount of colony-forming units per kidney of infected animals and was reversed by transformation of the cells with a plasmid carrying the *CaCLA4* gene (strain CLJ5CaCLA4) (Fig. 9). A histological examination revealed that kidneys from mice injected with either wild-type cells or cells deleted for one allele of *CaCLA4* were heavily infected with *C. albicans* cells that produced hyphae densely penetrating the animal tissue (Fig. 10, left panel), whereas kidneys from mice injected with cells deleted for both *CaCLA4* alleles contained small foci of aberrantly shaped cells that frequently carried multiple protrusions (Fig. 10, right panel). The morphologies of these cells were similar to those induced by serum under *in vitro* conditions. Thus, the function of CaCla4p is required for morphological switching of *C. albicans* under *in vitro* and *in vivo* conditions and for virulence.

Molecular cloning of the *CaCDC42* and *CaBEM1* genes

A *C. albicans* homolog of the *CaCDC42* gene was cloned by functional complementation of the temperature-sensitive growth defect of *S. cerevisiae* cells carrying the *cdc42-1^{ts}* mutation. The growth defect was fully complemented by plasmid YEp352-CaCDC42. The open reading frame of the *CaCDC42* gene is capable of encoding a protein of 191 amino acids with homology to the Rho-family of small G-proteins (Fig. 11). The highest homology is found with Cdc42p from *S. cerevisiae*.

A *C. albicans* homolog of the *CaBEM1* gene was cloned by functional complementation of the growth defect of *S. cerevisiae* cells deleted for the *BEM1* gene. This defect was fully complemented by plasmid YEp352-CaBEM1 carrying the *CaBEM1* gene. The open reading frame of the *CaBEM1* gene is capable of encoding a

protein of 635 amino acids with a domain structure characteristic of Bemlp (Fig. 12). CaBemlp contains two conserved SH3 domains which are most homologous to the SH3 domains of Bemlp, and also has homology to Bemlp outside of the SH3 domains.

Discussion

In *S. cerevisiae*, Ste20p fulfills multiple functions during mating (Leberer, E. et al. (1992) *EMBO J.* **11,** 4815-4824), pseudohyphae formation (Liu, H., Styles, C. & Fink, G. R. (1993) *Science* **262,** 1741-1744), invasive growth (Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8,** 2974-2985) and cytokinesis (Cvrckova, F. et al. (1995) *Genes Dev.* **9,** 1817-1830). *CST20* expression in *S. cerevisiae* fully complements these functions. Thus, Cst20p has the potential to fulfill similar functions in *C. albicans*.

The yeast-to-hyphal transition of *C. albicans* is a morphological change that can be triggered by a wide variety of factors. Carbohydrates, amino acids, salts, and serum have been described as inducers of germ tube formation, as have pH changes, temperature increases and starvation, but no single environmental factor could be defined as uniquely significant in stimulating the morphological switch. Hence *C. albicans* appears capable of responding to many divergent environmental signals. Disruption of both *CPH1* alleles, which encode a homolog of the *S. cerevisiae* Ste12p transcription factor (Liu, H. et al. (1994) *Science* **266,** 1723-1726), suppressed the lateral formation of mycelia from colonies grown on solid "Spider" medium, but did not block hyphal development in other media. We have shown that *C. albicans* mutant cells deleted for *CST20* display a similar phenotype, and that the effect of these mutations on hyphal development is dependent on the carbon source in which the cells were grown.

These observations are consistent with the idea that several signaling pathways can trigger morphogenesis in *C. albicans*. Furthermore, the behavior of *C. albicans* mutant strains deleted for either *CPH1* or *CST20* indicates that these pathways might operate independently to activate hyphal development under differing environmental conditions. *C. albicans* encounters a variety of different microenvironments during the development of superficial and systemic infections. Hence, the existence of parallel morphogenetic signaling pathways might provide a distinct advantage to this pathogen.

Our results indicate that the pathway controlled by Cst20p is not essential for virulence in a mouse model of systemic infections. It is not inconceivable that this pathway plays a role in other forms of infections, for example in the development of superficial infections of the mucosal epithelia (thrush). An as yet undefined role of Cst20p in pathogenicity outside of the Cst20 signaling pathway is suggested, however, by prolonged survival of mice infected with *cst20* deleted cells. It is unlikely that this effect is caused by defects in hyphal formation since a histological examination of infected kidneys revealed that the *CST20* deleted cells are not restricted in their capacity to form hyphae.

In *S. cerevisiae*, Cla4p plays a role in cytokinesis and shares with Ste20p an essential function for polarized growth during budding (Cvrckova, F. et al. (1995) *Genes Dev.* **9**, 1817-1830). Cla4p binds the Rho-like small G-protein Cdc42p (Cvrckova, F. et al. (1995) *Genes Dev.* **9**, 1817-1830) which is involved in controlling cell polarity during budding and in response to pheromone. Like Ste20p and the mammalian homolog p21-activated kinase ($p65^{PAK}$), Cla4p is able to phosphory-

late and activate myosin-I, a mechanism that may contribute to the organization of the actin cytoskeleton.

Our finding that *CaCLA4* expression in *S. cerevisiae* completely complements the Cla4p functions suggests that CaCla4p may have similar properties in *C. albicans*. Thus, CaCla4p may be required for myosin-I driven polarized growth during hyphal formation in a mechanism that may involve the *C. albicans* homolog of Cdc42p. Our complementation assays in *S. cerevisiae* suggest that CaCla4p may share an essential function with Cst20p, the *C. albicans* homolog of Ste20p (Figs. 6A and 6B). This notion suggests, together with our findings that null mutants of *CaCLA4* are completely non-pathogenic (Fig. 10) and null mutants of *CST20* are reduced in virulence (Figs. 6A and 6B), that CaCla4p and Cst20p, and proteins such as CaCdc42p and CaBem1p interacting with these protein kinases, may be valid targets for the development of antifungal agents.

The present invention will be more readily understood by referring to the following examples which are given to illustrate the invention rather than to limit its scope.

EXAMPLE I

Screening test for inhibitors of CaCla4p and Cst20p

An *in vitro* assay containing the proteins CaCla4p and/or Cst20p will be used to test compounds inhibiting their activity to render avirulent any fungi, which may be pathogenic.

The activity of the protein will be monitored to determine if the compounds tested do inhibit their biological activity, using myelin basic protein as a substrate.

In cases were a selective inhibition of CaCla4p and Cst20p and not to $p65^{PAK}$ would be desired, compounds testing positive for the inhibition of both

CaCla4p and Cst20p will be tested to determine if they also inhibit the protein $p65^{PAK}$. This would be useful in cases of pathogenic fungi infection such as for *C. albicans* were the fungi is to be rendered avirulent without affecting the normal protein of the patient $p65^{PAK}$.

In some cases of inflammation, it would be desirable to be provided with compounds inhibiting all three proteins, namely, CaCla4p, Cst20p and $p65^{PAK}$.

EXAMPLE II

Screening test for inhibitors of CaCla4p and CaCdc42p interactions

An *in vitro* assay containing the proteins CaCla4p and CaCdc42p will be used to test compounds inhibiting their interactions.

CaCla4p may be solid phase bound and CaCdc42p will be in suspension free to interact with CaCla4p. A labeled antibody specific to CaCdc42p will be added to the assay to determine the presence of CaCdc42p bound to CaCla4p. The compounds tested to inhibit the CaCdc42p-CaCla4p interactions, should when tested positive, cause only a minute quantity of CaCdc42p to bind to CaCla4p interactions.

The analogous *in vitro* assay will be used to test compounds that inhibit the interaction between Cst20p and CaCdc42p.

EXAMPLE III

Screening test for inhibitors of CaCla4p and CaBemlp interactions

An *in vitro* assay containing the proteins CaCla4p and CaBemlp will be used to test compounds inhibiting their interactions.

CaCla4p may be solid phase bound and CaBemlp will be in suspension free to interact with CaCla4p. A

labeled antibody specific to CaBem1p will be added to the assay to determine the presence of CaBem1p bound to CaCla4p. The compounds tested to inhibit the CaBem1p-CaCla4p interactions, should when tested positive, cause only a minute quantity of CaBem1p to bind to CaCla4p interactions.

The analogous *in vitro* assay will be used to test compounds that inhibit the interaction between Cst20p and CaBem1p.

EXAMPLE IV

A two-hybrid CaCdc42p and CaCla4p interaction system in a humanized *S. cerevisiae* strain

This screening assay is based on the assumption that the interaction of the small G-protein CaCdc42p with its cellular targets Cst20p and CaCla4p is essential for viability of C albicans cells. This essential function is reasonable to assume based on work that has been performed in *S. cerevisiae* (Leberer E. et al. (1997) Embo J. **16**, 83-97). The two hybrid interaction system will use green fluorescent protein fused to the *GAL1* promoter as a functional read out. This reporter gene will be integrated into a *S. cerevisiae* strain in which the *STE20* and *CLA4* genes have been replaced by the human homolog p65PAK. The *CaCDC42* gene will be fused to the DNA binding domain of *GAL4*, and the *CaCLA4* gene will be fused to the activation domain of *GAL4*. Interaction of the two proteins will cause green fluorescence. Whereas inhibitors of the interaction will suppress fluorescence.

Non-specific inhibitors of the two-hybrid interaction system will be excluded by performing a parallel screen with unrelated fusion proteins known to interact. Compounds of general toxicity or inhibitors of

the human homologs will also be excluded in this system because those compounds will not allow growth of the cells and therefore reduce the fluorescent readout in both parallel screens.

A two-hybrid yeast strain carrying the *GAL4-GFP* fusion gene is constructed. This strain will be deleted for the CLA4 gene using the TRP1 marker as described (Leberer E. et al. (1997) Embo J. **16**, 83-97). The *STE20* gene will be replaced by the human *PAK* gene as described above. To replace the *CDC42* gene by its human homolog, an integrating plasmid will be constructed carrying the *HsCDC42* gene fused to a *URA3* blaster gene and CDC42 flanking sequences. After linearization, the construct will be transformed into the *PAK* containing two-hybrid strain, and integrants will be selected on -ura medium. The *URA3* gene will then be looped out on FOA medium. The various gene disruptions and gene replacements will be verified by Southern blot analyses.

The two-hybrid vectors carrying the *CaCDC42* gene fused to the *GAL4*-DNA binding domain and the *CaCLA4* gene fused to the transcriptional activation domain of *GAL4* will be constructed by standard procedures. To facilitate the interaction of the two proteins, we will use site-directed mutagenesis to create a mutation in the CAAX-box domain of CaCDC42p to prevent isoprenylation and targeting of the fusion protein to the plasma membrane. We will evaluate and optimize the assay system and adapt the assay conditions to the scale used in microtiter plates for automated screening of compounds.

EXAMPLE V

Detection of the presence of *C. albicans* using probes

The sequences of either one of the genes *CaCLA4*, *CST20*, *CaCDC42* and *CaBEM1* may be used to derive probes for the detection of *C. albicans* using PCR techniques or hybridization assays.

EXAMPLE VI

Use of nucleotide sequences of *CaCLA4*, *CST20*, *CaCDC42* and *CaBEM1* to identify homologue from other fungi

The nucleotide sequences of *CaCLA4*, *CST20*, *CaCDC42* and *CaBEM1* may be used to identify and clone homologues from other fungi.

EXAMPLE VII

A *S. cerevisiae*-based screening system using CaSte20p and the pheromone signaling pathway as drug target

In this system, we will use green fluorescent protein (GFP) under transcriptional control of a pheromone inducible promoter (*FUS1*) as a read out. The pheromone signaling pathway and thereby the reporter gene will be induced with pheromone in two different strains. First, in a strain in which *STE20* is functionally replaced by the Ca*STE20* gene. And second, in a strain in which *STE20* is functionally replaced by the mammalian homolog *PAK*. Compounds that block the induction of the reporter gene in the Ca*STE20* strain but not in the *PAK* strain are expected to be specific inhibitors of the *C. albicans* kinase. This assay is very specific and is a positive selection of compounds that excludes the finding of compounds with inhibitory action against the mammalian homolog *PAK* or compounds of general toxicity.

The *FUS1* gene, including its promoter, will be isolated by the polymerase chain reaction (PCR) from

genomic DNA of *S. cerevisiae* and fused to the *GFP* gene from *Aequoria victoria* on a yeast expression plasmid. The function of the reporter gene will be analyzed after transformation of a *MAT*a yeast strain and induction with pheromone.

The *STE20* gene will be replaced in a supersensitive *sst1* yeast strain by the human *PAK* gene using homologous recombination. For this purpose, an integrating plasmid will be constructed carrying the *PAK* gene fused to a *URA3* blaster gene and *STE20* flanking sequences. The construct will be linearized and transformed into yeast, and integrants will be selected on -ura medium. The *URA3* gene will then be looped out on FOA medium to gain back the *ura3* marker. Correct integration of the *PAK* gene will be confirmed by Southern blot analysis.

The humanized strain will then be transformed with the *FUS1-GFP* reporter gene and analyzed for a functional signaling pathway by measuring green fluorescence after induction with pheromone. The assay system will be evaluated, optimized and adapted to the scale used in microtier plates.

EXAMPLE VIII

Fluorescence resonance energy transfer (FRET) as probe for protein-protein interactions

The engineering of different GFP mutants with altered fluorescence characteristics allows the use of fluorescence resonance energy transfer (FRET) to probe protein-protein interactions (Heim and Tsien (1996) Curr. Biol. **6**, 178-182). The FRET phenomenon consists in a fluorescence transfer between a donor and a receptor fluorochrome. If excitation and emission wavelengths are compatible, the FRET is easily measurable. The main parameter of the reaction is the distance

between donor and receptor, which must be in the range of nanometers. This is precisely the kind of values in protein-protein interactions.

We propose to develop a novel yeast assay system which uses FRET to measure the *in vivo* interaction between CaCdc42p and Cacla4p. The *CaCDC42* gene will be fused to a *GFP* mutant that acts as donor, and the *CaCLA4* gene will be fused to a mutant that acts as receptor. The yeast strain used as an expression system will be humanized as described in Example VII. Inhibitors of the interaction are expected to reduce energy transfer, and this reduction can be easily measured spectroscopically. The interaction of unrelated proteins known to interact will be used as a reference to exclude non-specific inhibitors of the assay system. Compounds inhibiting the interaction of the human homologs or of general toxicity will be excluded by inhibition of growth and therefore reduced fluorescence in both screens.

The *CaCDC42* gene will be fused to the gene encoding the GFP^{Y66H} mutant as donor, and the *CaCLA4* gene will be fused to the gene encoding the GFP^{S65T} mutant as receptor (Heim and Tsien (1996), Curr. Biol. **6**, 178-182). The constructs will then be transformed into the humanized yeast strain described in Example VII, and the FRET phenomenon will be analyzed in yeast cultures using fluorescence spectroscopy. The conditions for the assay will be worked out and optimized. We will adapt the assay conditions to the scale used in microtiter plates for automated screening.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following,

in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: National Research Council of Canada

(ii) TITLE OF THE INVENTION: *CANDIDA ALBICANS* PROTEINS ASSOCIATED WITH VIRULENCE AND HYPHAL FORMATION AND USES THEREOF

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SWABEY OGILVY RENAULT
(B) STREET: 1981 McGill College Ave. - Suite 1600
(C) CITY: Montréal
(D) STATE: QC
(E) COUNTRY: Canada
(F) ZIP: H3A 2Y3

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/029,458
(B) FILING DATE: 30-OCT-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Côté, France
(B) REGISTRATION NUMBER: 4166
(C) REFERENCE/DOCKET NUMBER: 2139-10PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 514 845-7126
(B) TELEFAX: 514-288-8389
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGGATCCAG ACCAACCACT CGAACTACT 29

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGGGATCCGA AGGTGAACCA CCATATTTG 29

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAAGATCTTG TAATCAATGT TCCCGTGGA 29

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAAGATCTCA TCGTGATATT AAATCCGAT 29

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4492 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic RNA

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 355...4044
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GTACCCACTT TACAATCACT TACAAGTCAA ATAATTACAA CTTGACAATC CTCACTTTAA     60
GTCTAACGTA TATACGCGTA CACCATCTTA TACTCCACAT ACATATTGGA TTCAATTTTT    120
ATTTTATTGT TTAGTTTATA TCCAACCACT GACAATTACC AATAGTTTTC AATTAATATT    180
CACAATTTAA CTATTTGTTT GACAGCTGAA AAGAGATAAA AAAAGAATCA AGTGCTATAA    240
CTCACAAGGG CTAGAAATAA GTTTGCAAAA AACAAGTTTT AAAAATAGTA ACTGCACTTT    300
TGTTGACTCT TTCACCTCCC CATTGAATTT AACTGAACAC AAATAAAGCC TATC ATG      357
                                                          Met
                                                           1

AGC ATA CTT TCA GAG AAC AAT CCT ACA CCA ACA TCA ATA ACA GAT CCA      405
Ser Ile Leu Ser Glu Asn Asn Pro Thr Pro Thr Ser Ile Thr Asp Pro
             5                  10                  15

AAT GAG TCT TCT CAT CTA CAC AAC CCA GAG TTA AAC TCT GGA ACG AGG      453
Asn Glu Ser Ser His Leu His Asn Pro Glu Leu Asn Ser Gly Thr Arg
        20                  25                  30

GTT GCT TCT GGA CCT GGA CCT GGA CCT GAA GTT GAA TCA ACA CCA CTA      501
Val Ala Ser Gly Pro Gly Pro Gly Pro Glu Val Glu Ser Thr Pro Leu
    35                  40                  45

GCA CCC CCA ACT GAG GTC ATG AAT ACT ACA TCA GCT AAT ACT TCT TCA      549
Ala Pro Pro Thr Glu Val Met Asn Thr Thr Ser Ala Asn Thr Ser Ser
50                  55                  60                  65

TTA AGT TTA GGG TCT CCA ATG CAC GAG AAA ATA AAA CAA TTT GAT CAA      597
Leu Ser Leu Gly Ser Pro Met His Glu Lys Ile Lys Gln Phe Asp Gln
                70                  75                  80

GAC GAG GTT GAC ACT GGG GAA ACT AAT GAT AGG ACT ATA GAA TCT GGA      645
Asp Glu Val Asp Thr Gly Glu Thr Asn Asp Arg Thr Ile Glu Ser Gly
            85                  90                  95

TCT AGT GAT ATT GAT GAT TCA CAA CAA TCA CAT AAC AAC AAC AAC AAC      693
Ser Ser Asp Ile Asp Asp Ser Gln Gln Ser His Asn Asn Asn Asn Asn
        100                 105                 110

AAC AAC AAC AAC AAC AAC GAG AGC AAT CCA GAA TCA AGT GAA GGC GAT      741
Asn Asn Asn Asn Asn Asn Glu Ser Asn Pro Glu Ser Ser Glu Gly Asp
    115                 120                 125

GAT GAA AAA ACC CAA GGA ATG CCT CCT CGA ATG CCA GGG ACA TTC AAT      789
Asp Glu Lys Thr Gln Gly Met Pro Pro Arg Met Pro Gly Thr Phe Asn
130                 135                 140                 145

GTG AAA GGT TTG CAC CAA GGG GAT GAT AGT GAC AAT GAA AAA CAG TAC      837
Val Lys Gly Leu His Gln Gly Asp Asp Ser Asp Asn Glu Lys Gln Tyr
                150                 155                 160

ACC GAG CTA ACT AAA TCA ATC AAT AAA CGT ACC AGT AAA GAT TCG TAT      885
Thr Glu Leu Thr Lys Ser Ile Asn Lys Arg Thr Ser Lys Asp Ser Tyr
            165                 170                 175
```

```
TCT CCT GGC ACA CTT GAA AGT CCC GGT ACT CTT AAT GCA TTG GAA ACA      933
Ser Pro Gly Thr Leu Glu Ser Pro Gly Thr Leu Asn Ala Leu Glu Thr
        180                 185                 190

AAT AAT GTC TCA CCA GCA GTT ATA GAG GAA GAA CAA CAT ACA CTG TCT      981
Asn Asn Val Ser Pro Ala Val Ile Glu Glu Glu Gln His Thr Leu Ser
    195                 200                 205

TTG GAA GAT TTG TCA TTG TCC TTA CAA CAC CAA AAT GAA AAT GCA AGA     1029
Leu Glu Asp Leu Ser Leu Ser Leu Gln His Gln Asn Glu Asn Ala Arg
210                 215                 220                 225

TTA TCT GCA CCC CGC AGT GCA CCG CCA CAG GTT CCG ACT TCA AAG ACA     1077
Leu Ser Ala Pro Arg Ser Ala Pro Pro Gln Val Pro Thr Ser Lys Thr
                230                 235                 240

TCG TCA TTT CAC GAT ATG AGT CTG GTT ATA TCT TCA TCA ACT TCT GTG     1125
Ser Ser Phe His Asp Met Ser Leu Val Ile Ser Ser Ser Thr Ser Val
            245                 250                 255

CAT AAG ATA CCA TCA AAT CCA ACT TCA ACT CGA GGT TCT CAT TTA TCA     1173
His Lys Ile Pro Ser Asn Pro Thr Ser Thr Arg Gly Ser His Leu Ser
        260                 265                 270

AGT TAC AAA TCT ACA TTG GAC CCT GGG AAA CCT GCA CAA GCA GCA GCA     1221
Ser Tyr Lys Ser Thr Leu Asp Pro Gly Lys Pro Ala Gln Ala Ala Ala
    275                 280                 285

CCA CCA CCA CCA GAA ATA GAC ATT GAC AAT TTA TTA ACC AAA AGT GAA     1269
Pro Pro Pro Pro Glu Ile Asp Ile Asp Asn Leu Leu Thr Lys Ser Glu
290                 295                 300                 305

TTG GAT CTG GAA ACA GAC ACA TTG AGT AGT GCC ACA AAT TCT CCA AAC     1317
Leu Asp Leu Glu Thr Asp Thr Leu Ser Ser Ala Thr Asn Ser Pro Asn
                310                 315                 320

CTT TTA AGA AAT GAT ACT TTA CAA GGA ATT CCA ACA AGA GAT GAC GAA     1365
Leu Leu Arg Asn Asp Thr Leu Gln Gly Ile Pro Thr Arg Asp Asp Glu
            325                 330                 335

AAT ATT GAT GAC CTG CCC CGT CAA CTA TCA CAA AAT ACT AGT GCG ACG     1413
Asn Ile Asp Asp Leu Pro Arg Gln Leu Ser Gln Asn Thr Ser Ala Thr
        340                 345                 350

TCA AGA AAT ACT TCG GGA ACA TCG ACT TCT ACA GTG GTG AAA AAT TCA     1461
Ser Arg Asn Thr Ser Gly Thr Ser Thr Ser Thr Val Val Lys Asn Ser
    355                 360                 365

AGA TCT GGT ACG TCA AAA TCA ACC TCA ACC TCA ACT GCT CAT AAC CAA     1509
Arg Ser Gly Thr Ser Lys Ser Thr Ser Thr Ser Thr Ala His Asn Gln
370                 375                 380                 385

ACA GCA GCA ATT ACT CCT ATA ATC CCG AGT CAC AAC AAG TTT CAT CAA     1557
Thr Ala Ala Ile Thr Pro Ile Ile Pro Ser His Asn Lys Phe His Gln
                390                 395                 400

CAA GTG ATA AAT ACC AAT GCA ACA AAT AGT TCA TCT TCA CTA GAA CCA     1605
Gln Val Ile Asn Thr Asn Ala Thr Asn Ser Ser Ser Ser Leu Glu Pro
            405                 410                 415
```

```
TTG GGG GTT GGC ATA AAT TCA AAT CTG TCT CCT AAA AGT GGG AAA AAG    1653
Leu Gly Val Gly Ile Asn Ser Asn Leu Ser Pro Lys Ser Gly Lys Lys
        420                 425                 430

CGG AAA AGT GGA AGT AAA GTC CGA GGT GTG TTT TCG TCA ATG TTT GGG    1701
Arg Lys Ser Gly Ser Lys Val Arg Gly Val Phe Ser Ser Met Phe Gly
    435                 440                 445

AAA AAC AAG TCA ACG TCA TCA TCG TCG TCT TCA AAC TCA GGT CTG AAT    1749
Lys Asn Lys Ser Thr Ser Ser Ser Ser Ser Ser Asn Ser Gly Leu Asn
450                 455                 460                 465

AGC CAC TCA CAG GAA GTC AAT ATT AAG ATC AGT ACT CCA TTC AAT GCC    1797
Ser His Ser Gln Glu Val Asn Ile Lys Ile Ser Thr Pro Phe Asn Ala
                470                 475                 480

AAG CAC CTT GCC CAT GTG GGC ATT GAT GAT AAT GGT TCA TAC ACC GGT    1845
Lys His Leu Ala His Val Gly Ile Asp Asp Asn Gly Ser Tyr Thr Gly
            485                 490                 495

TTG CCA ATA GAG TGG GAA AGA TTA TTA TCT GCT AGT GGT ATT ACC AAG    1893
Leu Pro Ile Glu Trp Glu Arg Leu Leu Ser Ala Ser Gly Ile Thr Lys
        500                 505                 510

AAG GAA CAA CAA CAG CAC CCA CAA GCA GTG ATG GAT ATA GTG GCG TTT    1941
Lys Glu Gln Gln Gln His Pro Gln Ala Val Met Asp Ile Val Ala Phe
    515                 520                 525

TAT CAA GAT ACA AGT GAA AAC CCT GAT GAC GCT GCA TTT AAA AAG TTT    1989
Tyr Gln Asp Thr Ser Glu Asn Pro Asp Asp Ala Ala Phe Lys Lys Phe
530                 535                 540                 545

CAT TTT GAT AAT AAT AAA AGT AGT TCG AGT GGT TGG TCT AAT GAA AAT    2037
His Phe Asp Asn Asn Lys Ser Ser Ser Ser Gly Trp Ser Asn Glu Asn
                550                 555                 560

ACT CCA CCA GCA ACA CCG GGT GGG AGT AAC AGT GGC AGT GGC AGT GGT    2085
Thr Pro Pro Ala Thr Pro Gly Gly Ser Asn Ser Gly Ser Gly Ser Gly
            565                 570                 575

GGC GGT GGC GCT CCT TCA AGT CCC CAT CGT ACA CCT CCT TCA TCG ATC    2133
Gly Gly Gly Ala Pro Ser Ser Pro His Arg Thr Pro Pro Ser Ser Ile
        580                 585                 590

ATT GAA AAA AAC AAC GTT GAA CAA AAA GTG ATT ACC CCA TCT CAG TCA    2181
Ile Glu Lys Asn Asn Val Glu Gln Lys Val Ile Thr Pro Ser Gln Ser
    595                 600                 605

ATG CCA ACA AAG ACA GAG AGT AAA CAG CTG GAA AAC CAG CAC CCA CAT    2229
Met Pro Thr Lys Thr Glu Ser Lys Gln Leu Glu Asn Gln His Pro His
610                 615                 620                 625

GAA GAT AAT GCT ACT CAG TAT ACA CCA AGA ACA CCA ACA TCC CAT GTA    2277
Glu Asp Asn Ala Thr Gln Tyr Thr Pro Arg Thr Pro Thr Ser His Val
                630                 635                 640
```

```
CAA GAG GGT CAA TTT ATT CCA AGT AGA CCA GCT CCG AAA CCA CCA TCA     2325
Gln Glu Gly Gln Phe Ile Pro Ser Arg Pro Ala Pro Lys Pro Pro Ser
            645                 650                 655

ACA CCG CTT TCT TCC ATG AGT GTG TCA CAT AAA ACA CCT TCT TCG CAA     2373
Thr Pro Leu Ser Ser Met Ser Val Ser His Lys Thr Pro Ser Ser Gln
        660                 665                 670

TCA TTA CCA AGG AGT GAT TCA CAA TCC GAT ATT CGT TCT TCA ACC CCT     2421
Ser Leu Pro Arg Ser Asp Ser Gln Ser Asp Ile Arg Ser Ser Thr Pro
    675                 680                 685

AAA TCA CAT CAA GAT GTT TCG CCA AGC AAG ATC AAA ATT CGT TCA ATT     2469
Lys Ser His Gln Asp Val Ser Pro Ser Lys Ile Lys Ile Arg Ser Ile
690                 695                 700                 705

TCG TCA AAA TCA TTA AAG TCA ATG CGG TCT AGA AAA AGT GGG GAT AAG     2517
Ser Ser Lys Ser Leu Lys Ser Met Arg Ser Arg Lys Ser Gly Asp Lys
                710                 715                 720

TTT ACT CAT ATT GCA CCT GCT CCT CCA CCA CCA TCA TTA CCT TCA ATT     2565
Phe Thr His Ile Ala Pro Ala Pro Pro Pro Pro Ser Leu Pro Ser Ile
            725                 730                 735

CCT AAA TCA AAG TCG CAT TCG GCA TCT TTG TCA AGT CAA TTG AGA CCA     2613
Pro Lys Ser Lys Ser His Ser Ala Ser Leu Ser Ser Gln Leu Arg Pro
        740                 745                 750

GCA ACA AAT GGA TCA ACA ACT GCC CCT ATT CCA GCA AGT GCC GCG TTT     2661
Ala Thr Asn Gly Ser Thr Thr Ala Pro Ile Pro Ala Ser Ala Ala Phe
    755                 760                 765

GGT GGT GAG AAT AAT GCT TTA CCA AAA CAA AGA ATA AAT GAG TTC AAG     2709
Gly Gly Glu Asn Asn Ala Leu Pro Lys Gln Arg Ile Asn Glu Phe Lys
770                 775                 780                 785

GCT CAT AGA GCA CCT CCA CCA CCT CCA CTG GCA CCA CCT GCA CCA CCT     2757
Ala His Arg Ala Pro Pro Pro Pro Pro Leu Ala Pro Pro Ala Pro Pro
                790                 795                 800

GTG CCT CCT GCT CCA CCA GCC AAT TTA TTA TCG GAA CAG ACT TCT GAG     2805
Val Pro Pro Ala Pro Pro Ala Asn Leu Leu Ser Glu Gln Thr Ser Glu
            805                 810                 815

ATA CCT CAA CAA CGT ACT GCT CCT CTG CAA GCA TTA GCT GAT GTT ACT     2853
Ile Pro Gln Gln Arg Thr Ala Pro Leu Gln Ala Leu Ala Asp Val Thr
        820                 825                 830

GCC CCA ACT AAT ATT TAT GAA ATT CAA CAA ACT AAA TAT CAG GAA GCA     2901
Ala Pro Thr Asn Ile Tyr Glu Ile Gln Gln Thr Lys Tyr Gln Glu Ala
    835                 840                 845

CAA CAG AAA TTA CGT GAG AAG AAG GCT AGA GAA CTT GAA GAA ATA CAA     2949
Gln Gln Lys Leu Arg Glu Lys Lys Ala Arg Glu Leu Glu Glu Ile Gln
850                 855                 860                 865

AGA CTA CGA GAG AAG AAT GAA AGA CAA AAT AGA CAA CAG GAG ACT GGG     2997
Arg Leu Arg Glu Lys Asn Glu Arg Gln Asn Arg Gln Gln Glu Thr Gly
                870                 875                 880
```

```
CAA AAT AAT GCT GAC ACG GCT AGC GGT GGT AGT AAT ATT GCT CCA CCA    3045
Gln Asn Asn Ala Asp Thr Ala Ser Gly Gly Ser Asn Ile Ala Pro Pro
            885                 890                 895

GTA CCT GTA CCA AAT AAA AAA CCG CCT TCT GGA TCT GGT GGT GGC CGT    3093
Val Pro Val Pro Asn Lys Lys Pro Pro Ser Gly Ser Gly Gly Gly Arg
        900                 905                 910

GAT GCC AAA CAA GCA GCT TTG ATA GCC CAA AAG AAA CGA GAA GAA AAG    3141
Asp Ala Lys Gln Ala Ala Leu Ile Ala Gln Lys Lys Arg Glu Glu Lys
    915                 920                 925

AAA CGT AAA AAC TTA CAA ATT ATT GCC AAA TTA AAG ACA ATT TGT AAT    3189
Lys Arg Lys Asn Leu Gln Ile Ile Ala Lys Leu Lys Thr Ile Cys Asn
930                 935                 940                 945

CCT GGA GAT CCA AAT GAA TTA TAT GTT GAT TTA GTT AAA ATT GGT CAA    3237
Pro Gly Asp Pro Asn Glu Leu Tyr Val Asp Leu Val Lys Ile Gly Gln
                950                 955                 960

GGT GCC TCC GGT GGA GTT TTC CTT GCT CAT GAT GTT CGT GAT AAA TCC    3285
Gly Ala Ser Gly Gly Val Phe Leu Ala His Asp Val Arg Asp Lys Ser
            965                 970                 975

AAT ATT GTT GCC ATA AAA CAA ATG AAT TTA GAA CAA CAA CCT AAA AAA    3333
Asn Ile Val Ala Ile Lys Gln Met Asn Leu Glu Gln Gln Pro Lys Lys
        980                 985                 990

GAA TTA ATT ATT AAT GAA ATT TTG GTT ATG AAA GGT AGT CTG CAT CCC    3381
Glu Leu Ile Ile Asn Glu Ile Leu Val Met Lys Gly Ser Leu His Pro
    995                1000                1005

AAT ATT GTC AAT TTT ATT GAT TCA TAT CTT TTA AAA GGT GAT TTA TGG    3429
Asn Ile Val Asn Phe Ile Asp Ser Tyr Leu Leu Lys Gly Asp Leu Trp
1010                1015                1020                1025

GTG ATT ATG GAA TAT ATG GAA GGT GGA TCC CTT ACC GAT ATA GTG ACT    3477
Val Ile Met Glu Tyr Met Glu Gly Gly Ser Leu Thr Asp Ile Val Thr
                1030                1035                1040

CAT AGT GTT ATG ACC GAA GGT CAA ATT GGA GTT GTA TGT CGT GAA ACT    3525
His Ser Val Met Thr Glu Gly Gln Ile Gly Val Val Cys Arg Glu Thr
            1045                1050                1055

TTG AAA GGT CTT AAA TTT TTA CAT TCC AAA GGG GTT ATC CAT CGT GAT    3573
Leu Lys Gly Leu Lys Phe Leu His Ser Lys Gly Val Ile His Arg Asp
        1060                1065                1070

ATT AAA TCC GAT AAT ATT TTA TTA AAT ATG GAT GGT AAC ATC AAG ATC    3621
Ile Lys Ser Asp Asn Ile Leu Leu Asn Met Asp Gly Asn Ile Lys Ile
    1075                1080                1085

ACT GAT TTT GGG TTT TGT GCT CAA ATC AAT GAA ATC AAT CTG AAA CGT    3669
Thr Asp Phe Gly Phe Cys Ala Gln Ile Asn Glu Ile Asn Leu Lys Arg
1090                1095                1100                1105
```

```
ATC ACT ATG GTG GGT ACA CCA TAT TGG ATG GCA CCA GAA ATT GTT TCA    3717
Ile Thr Met Val Gly Thr Pro Tyr Trp Met Ala Pro Glu Ile Val Ser
                1110                1115                1120

CGT AAA GAG TAT GGT CCA AAA GTT GAT GTT TGG TCA TTA GGT ATC ATG    3765
Arg Lys Glu Tyr Gly Pro Lys Val Asp Val Trp Ser Leu Gly Ile Met
            1125                1130                1135

ATT ATA GAA ATG TTA GAA GGT GAA CCA CCA TAT TTG AAT GAA ACT CCA    3813
Ile Ile Glu Met Leu Glu Gly Glu Pro Pro Tyr Leu Asn Glu Thr Pro
        1140                1145                1150

TTG AGG GCA TTA TAT CTT ATT GCA ACT AAT GGT ACA CCA AAA TTA AAA    3861
Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly Thr Pro Lys Leu Lys
    1155                1160                1165

GAT CCT GAA TCT TTA AGT TAT GAT ATT AGA AAA TTT TTG GCA TGG TGT    3909
Asp Pro Glu Ser Leu Ser Tyr Asp Ile Arg Lys Phe Leu Ala Trp Cys
1170                1175                1180                1185

TTA CAA GTT GAC TTT AAT AAA AGA GCT GAT GCT GAT GAA TTA TTA CAT    3957
Leu Gln Val Asp Phe Asn Lys Arg Ala Asp Ala Asp Glu Leu Leu His
                1190                1195                1200

GAT AAT TTT ATT ACT GAA TGT GAT GAT GTA TCG TCG TTA AGT CCA TTA    4005
Asp Asn Phe Ile Thr Glu Cys Asp Asp Val Ser Ser Leu Ser Pro Leu
            1205                1210                1215

GTG AAA ATT GCT CGA TTG AAA AAA ATG AGT GAA TCT GAT TAATGAATGG TG  4056
Val Lys Ile Ala Arg Leu Lys Lys Met Ser Glu Ser Asp
        1220                1225                1230

GAGTTATCCT AGAAATAAAA ACATTTAAAA AAAAAGAAGA AGAACAACAA GAACCCTAAA   4116
TTCTACTGCT GTCAATATAT TGGCTAATTT CCATTCTCGT TTCTATTTCT ATTTCGTTTT   4176
TATTCTTTGA ATTATTATTG TTAGTGGTAG AGATTTTTAC TAGTATATTT TTTTTATTCA   4236
TTTTTATATT TGTATTTATA TATATATTTT TCATTTAGTA TTTACTTACA CTGCAGTATC   4296
TTTCTTTTCT GTGTAGATGA TATGTAGTAA TAAGTTAACT TGTTCAAGAC AGTGAATGGA   4356
AATATATTAT AGCTTGACTA TATAAGGTGG AGAGCTGTAA TTGGCTTTCC GTATAGAAAA   4416
GTCTTGAACA AACGTTACCA GATTTCTGCT ATTCTTATTT GGTACGATTC GGGCGTATGA   4476
TAGGTTTATT GAGCTC                                                   4492
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 1230 amino acids
  (B) TYPE: amino acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ser Ile Leu Ser Glu Asn Asn Pro Thr Pro Thr Ser Ile Thr Asp
 1               5                  10                  15
Pro Asn Glu Ser Ser His Leu His Asn Pro Glu Leu Asn Ser Gly Thr
            20                  25                  30
```

```
Arg Val Ala Ser Gly Pro Gly Pro Gly Pro Glu Val Glu Ser Thr Pro
        35                  40                  45
Leu Ala Pro Pro Thr Glu Val Met Asn Thr Thr Ser Ala Asn Thr Ser
    50                  55                  60
Ser Leu Ser Leu Gly Ser Pro Met His Glu Lys Ile Lys Gln Phe Asp
65                  70                  75                  80
Gln Asp Glu Val Asp Thr Gly Glu Thr Asn Asp Arg Thr Ile Glu Ser
                85                  90                  95
Gly Ser Ser Asp Ile Asp Asp Ser Gln Gln Ser His Asn Asn Asn Asn
            100                 105                 110
Asn Asn Asn Asn Asn Asn Asn Glu Ser Asn Pro Glu Ser Ser Glu Gly
        115                 120                 125
Asp Asp Glu Lys Thr Gln Gly Met Pro Pro Arg Met Pro Gly Thr Phe
    130                 135                 140
Asn Val Lys Gly Leu His Gln Gly Asp Asp Ser Asp Asn Glu Lys Gln
145                 150                 155                 160
Tyr Thr Glu Leu Thr Lys Ser Ile Asn Lys Arg Thr Ser Lys Asp Ser
                165                 170                 175
Tyr Ser Pro Gly Thr Leu Glu Ser Pro Gly Thr Leu Asn Ala Leu Glu
            180                 185                 190
Thr Asn Asn Val Ser Pro Ala Val Ile Glu Glu Glu Gln His Thr Leu
        195                 200                 205
Ser Leu Glu Asp Leu Ser Leu Ser Leu Gln His Gln Asn Glu Asn Ala
    210                 215                 220
Arg Leu Ser Ala Pro Arg Ser Ala Pro Pro Gln Val Pro Thr Ser Lys
225                 230                 235                 240
Thr Ser Ser Phe His Asp Met Ser Leu Val Ile Ser Ser Ser Thr Ser
                245                 250                 255
Val His Lys Ile Pro Ser Asn Pro Thr Ser Thr Arg Gly Ser His Leu
            260                 265                 270
Ser Ser Tyr Lys Ser Thr Leu Asp Pro Gly Lys Pro Ala Gln Ala Ala
        275                 280                 285
Ala Pro Pro Pro Pro Glu Ile Asp Ile Asp Asn Leu Leu Thr Lys Ser
    290                 295                 300
Glu Leu Asp Leu Glu Thr Asp Thr Leu Ser Ser Ala Thr Asn Ser Pro
305                 310                 315                 320
Asn Leu Leu Arg Asn Asp Thr Leu Gln Gly Ile Pro Thr Arg Asp Asp
                325                 330                 335
Glu Asn Ile Asp Asp Leu Pro Arg Gln Leu Ser Gln Asn Thr Ser Ala
            340                 345                 350
Thr Ser Arg Asn Thr Ser Gly Thr Ser Thr Ser Thr Val Val Lys Asn
        355                 360                 365
Ser Arg Ser Gly Thr Ser Lys Ser Thr Ser Thr Ser Thr Ala His Asn
    370                 375                 380
Gln Thr Ala Ala Ile Thr Pro Ile Ile Pro Ser His Asn Lys Phe His
385                 390                 395                 400
Gln Gln Val Ile Asn Thr Asn Ala Thr Asn Ser Ser Ser Ser Leu Glu
                405                 410                 415
Pro Leu Gly Val Gly Ile Asn Ser Asn Leu Ser Pro Lys Ser Gly Lys
            420                 425                 430
Lys Arg Lys Ser Gly Ser Lys Val Arg Gly Val Phe Ser Ser Met Phe
        435                 440                 445
Gly Lys Asn Lys Ser Thr Ser Ser Ser Ser Ser Ser Asn Ser Gly Leu
    450                 455                 460
Asn Ser His Ser Gln Glu Val Asn Ile Lys Ile Ser Thr Pro Phe Asn
465                 470                 475                 480
Ala Lys His Leu Ala His Val Gly Ile Asp Asp Asn Gly Ser Tyr Thr
                485                 490                 495
```

```
Gly Leu Pro Ile Glu Trp Glu Arg Leu Leu Ser Ala Ser Gly Ile Thr
                500                 505                 510
Lys Lys Glu Gln Gln Gln His Pro Gln Ala Val Met Asp Ile Val Ala
            515                 520                 525
Phe Tyr Gln Asp Thr Ser Glu Asn Pro Asp Asp Ala Ala Phe Lys Lys
        530                 535                 540
Phe His Phe Asp Asn Asn Lys Ser Ser Ser Ser Gly Trp Ser Asn Glu
545                 550                 555                 560
Asn Thr Pro Pro Ala Thr Pro Gly Gly Ser Asn Ser Gly Ser Gly Ser
                565                 570                 575
Gly Gly Gly Gly Ala Pro Ser Ser Pro His Arg Thr Pro Pro Ser Ser
            580                 585                 590
Ile Ile Glu Lys Asn Asn Val Glu Gln Lys Val Ile Thr Pro Ser Gln
        595                 600                 605
Ser Met Pro Thr Lys Thr Glu Ser Lys Gln Leu Glu Asn Gln His Pro
    610                 615                 620
His Glu Asp Asn Ala Thr Gln Tyr Thr Pro Arg Thr Pro Thr Ser His
625                 630                 635                 640
Val Gln Glu Gly Gln Phe Ile Pro Ser Arg Pro Ala Pro Lys Pro Pro
                645                 650                 655
Ser Thr Pro Leu Ser Ser Met Ser Val Ser His Lys Thr Pro Ser Ser
            660                 665                 670
Gln Ser Leu Pro Arg Ser Asp Ser Gln Ser Asp Ile Arg Ser Ser Thr
        675                 680                 685
Pro Lys Ser His Gln Asp Val Ser Pro Ser Lys Ile Lys Ile Arg Ser
    690                 695                 700
Ile Ser Ser Lys Ser Leu Lys Ser Met Arg Ser Arg Lys Ser Gly Asp
705                 710                 715                 720
Lys Phe Thr His Ile Ala Pro Ala Pro Pro Pro Pro Ser Leu Pro Ser
                725                 730                 735
Ile Pro Lys Ser Lys Ser His Ser Ala Ser Leu Ser Ser Gln Leu Arg
            740                 745                 750
Pro Ala Thr Asn Gly Ser Thr Thr Ala Pro Ile Pro Ala Ser Ala Ala
        755                 760                 765
Phe Gly Gly Glu Asn Asn Ala Leu Pro Lys Gln Arg Ile Asn Glu Phe
    770                 775                 780
Lys Ala His Arg Ala Pro Pro Pro Pro Pro Leu Ala Pro Pro Ala Pro
785                 790                 795                 800
Pro Val Pro Pro Ala Pro Pro Ala Asn Leu Leu Ser Glu Gln Thr Ser
                805                 810                 815
Glu Ile Pro Gln Gln Arg Thr Ala Pro Leu Gln Ala Leu Ala Asp Val
            820                 825                 830
Thr Ala Pro Thr Asn Ile Tyr Glu Ile Gln Gln Thr Lys Tyr Gln Glu
        835                 840                 845
Ala Gln Gln Lys Leu Arg Glu Lys Lys Ala Arg Glu Leu Glu Glu Ile
    850                 855                 860
Gln Arg Leu Arg Glu Lys Asn Glu Arg Gln Asn Arg Gln Gln Glu Thr
865                 870                 875                 880
Gly Gln Asn Asn Ala Asp Thr Ala Ser Gly Gly Ser Asn Ile Ala Pro
                885                 890                 895
Pro Val Pro Val Pro Asn Lys Lys Pro Pro Ser Gly Ser Gly Gly Gly
            900                 905                 910
Arg Asp Ala Lys Gln Ala Ala Leu Ile Ala Gln Lys Lys Arg Glu Glu
        915                 920                 925
Lys Lys Arg Lys Asn Leu Gln Ile Ile Ala Lys Leu Lys Thr Ile Cys
    930                 935                 940
Asn Pro Gly Asp Pro Asn Glu Leu Tyr Val Asp Leu Val Lys Ile Gly
945                 950                 955                 960
```

```
Gln Gly Ala Ser Gly Gly Val Phe Leu Ala His Asp Val Arg Asp Lys
                965                 970                 975
Ser Asn Ile Val Ala Ile Lys Gln Met Asn Leu Glu Gln Gln Pro Lys
            980                 985                 990
Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Lys Gly Ser Leu His
        995                1000                1005
Pro Asn Ile Val Asn Phe Ile Asp Ser Tyr Leu Leu Lys Gly Asp Leu
    1010                1015                1020
Trp Val Ile Met Glu Tyr Met Glu Gly Gly Ser Leu Thr Asp Ile Val
1025                1030                1035                1040
Thr His Ser Val Met Thr Glu Gly Gln Ile Gly Val Val Cys Arg Glu
                1045                1050                1055
Thr Leu Lys Gly Leu Lys Phe Leu His Ser Lys Gly Val Ile His Arg
            1060                1065                1070
Asp Ile Lys Ser Asp Asn Ile Leu Leu Asn Met Asp Gly Asn Ile Lys
        1075                1080                1085
Ile Thr Asp Phe Gly Phe Cys Ala Gln Ile Asn Glu Ile Asn Leu Lys
    1090                1095                1100
Arg Ile Thr Met Val Gly Thr Pro Tyr Trp Met Ala Pro Glu Ile Val
1105                1110                1115                1120
Ser Arg Lys Glu Tyr Gly Pro Lys Val Asp Val Trp Ser Leu Gly Ile
                1125                1130                1135
Met Ile Ile Glu Met Leu Glu Gly Glu Pro Pro Tyr Leu Asn Glu Thr
            1140                1145                1150
Pro Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly Thr Pro Lys Leu
        1155                1160                1165
Lys Asp Pro Glu Ser Leu Ser Tyr Asp Ile Arg Lys Phe Leu Ala Trp
    1170                1175                1180
Cys Leu Gln Val Asp Phe Asn Lys Arg Ala Asp Ala Asp Glu Leu Leu
 185                1190                1195                1200
His Asp Asn Phe Ile Thr Glu Cys Asp Asp Val Ser Ser Leu Ser Pro
                1205                1210                1215
Leu Val Lys Ile Ala Arg Leu Lys Lys Met Ser Glu Ser Asp
            1220                1225                1230
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 432...3344
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GAATTCTTTT TAGAAGAGAA AGAAAAAATT CCCAAAAAAA AAAGATTTCA TTTAATTCCA     60
CGGGAACATT GATTACAACC ACGTCAACAG TTTCCCTTTT ATATTGAAAT CAACATTCAA    120
TTTTGTCTTT TTTTTTTTTT CATTGATTTT TCCCCAATCT TTTTATCTTC ATATTAATAT    180
TGGATATCAA TTACTAATAC TGTCAGGGAT AGTTTAGTAA ATATTTACAT TCTCCATTCA    240
ATCCTAAATT TTTTTTTATA TAGCTAGTTT TTGGTTGAAA AAAAAAAAAT AGGGGGAAGG    300
AAGTTTTTTT TTCTATTTAT TTAATTGTTT TGATTCCAAC CATATTGTAT ATTTGTCTTG    360
```

```
TCAGTTATAT TACTTTCTTG TTACTTAATT ATTAATTATT TGCTATATTA TTGAATTGAA    420
TCCTCAAAAG A ATG ACA AGT ATT TAT ACA TCA GAT TTG AAA AAC CAT AGA    470
             Met Thr Ser Ile Tyr Thr Ser Asp Leu Lys Asn His Arg
              1               5                  10

CGT GCG CCA CCT CCA CCA AAT GGG GCA GCT GGC TCT GGC TCA GGT TCT     518
Arg Ala Pro Pro Pro Pro Asn Gly Ala Ala Gly Ser Gly Ser Gly Ser
    15                  20                  25

GGC TCA GGT TCT GGT TCT GGT TCT GGC AGT TTG GCT AAT ATT GTT ACC     566
Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Leu Ala Asn Ile Val Thr
30                  35                  40                  45

AGT TCT AAT AGT CTT GGC GTA ACA GCA AAT CAA ACC AAA CCT ATT CAA     614
Ser Ser Asn Ser Leu Gly Val Thr Ala Asn Gln Thr Lys Pro Ile Gln
                50                  55                  60

TTA AAT ATA AAT TCT AGC AAA CGT CAA TCA GGT TGG GTT CAT GTT AAA     662
Leu Asn Ile Asn Ser Ser Lys Arg Gln Ser Gly Trp Val His Val Lys
            65                  70                  75

GAT GAT GGT ATT TTC ACA TCA TTT AGA TGG AAC AAA CGG TTT ATG GTT     710
Asp Asp Gly Ile Phe Thr Ser Phe Arg Trp Asn Lys Arg Phe Met Val
        80                  85                  90

ATT AAT GAT AAA ACT TTA AAC TTT TAT AAA CAA GAA CCA TAT TCT AGT     758
Ile Asn Asp Lys Thr Leu Asn Phe Tyr Lys Gln Glu Pro Tyr Ser Ser
    95                  100                 105

GAT GGT AAT TCC AAT TCT AAT ACC CCT GAT TTA TCA TTC CCA CTA TAT     806
Asp Gly Asn Ser Asn Ser Asn Thr Pro Asp Leu Ser Phe Pro Leu Tyr
110                 115                 120                 125

TTA ATT AAT AAT ATT AAT TTG AAA CCA AAC TCC GGG TAT AGC AAA ACT     854
Leu Ile Asn Asn Ile Asn Leu Lys Pro Asn Ser Gly Tyr Ser Lys Thr
                130                 135                 140

TCA CAA TCA TTT GAA ATT GTT CCC AAA AAC AAT AAT AAA TCA ATT TTG     902
Ser Gln Ser Phe Glu Ile Val Pro Lys Asn Asn Asn Lys Ser Ile Leu
            145                 150                 155

ATT TCT GTT AAA ACC AAT AAT GAT TAT TTG GAT TGG CTA GAT GCA TTC     950
Ile Ser Val Lys Thr Asn Asn Asp Tyr Leu Asp Trp Leu Asp Ala Phe
        160                 165                 170

ACC ACA AAA TGT CCT TTA GTA CAA ATT GGT GAA AAT AAT AGT GGT GTA     998
Thr Thr Lys Cys Pro Leu Val Gln Ile Gly Glu Asn Asn Ser Gly Val
    175                 180                 185

TCA AGT AGT CAC CCT CAT TTA CAA ATT CAA CAT TTA ACC AAT GGT TCC    1046
Ser Ser Ser His Pro His Leu Gln Ile Gln His Leu Thr Asn Gly Ser
190                 195                 200                 205

TTG AAC GGC AAC TCA TCT TCA TCA CCA ACA TCT GGA TTA TTA TCT TCT    1094
Leu Asn Gly Asn Ser Ser Ser Ser Pro Thr Ser Gly Leu Leu Ser Ser
                210                 215                 220
```

```
TCA GTG CTA ACT GGA GGT AAT TCT GGC GTT TCT GGT CCT ATT AAT TTC    1142
Ser Val Leu Thr Gly Gly Asn Ser Gly Val Ser Gly Pro Ile Asn Phe
            225                 230                 235

ACT CAT AAA GTA CAC GTG GGA TTT GAT CCT GCC AGT GGT AAT TTT ACT    1190
Thr His Lys Val His Val Gly Phe Asp Pro Ala Ser Gly Asn Phe Thr
        240                 245                 250

GGA TTA CCA GAC ACT TGG AAA AGT TTA TTA CAA CAT TCG AAA ATC ACT    1238
Gly Leu Pro Asp Thr Trp Lys Ser Leu Leu Gln His Ser Lys Ile Thr
    255                 260                 265

AAT GAG GAT TGG AAA AAA GAT CCT GTT GCT GTT ATT GAA GTT TTA GAA    1286
Asn Glu Asp Trp Lys Lys Asp Pro Val Ala Val Ile Glu Val Leu Glu
270                 275                 280                 285

TTT TAT TCC GAT ATA AAT GGA GGT AAT TCA GCT GCT GGA ACT CCA ATT    1334
Phe Tyr Ser Asp Ile Asn Gly Gly Asn Ser Ala Ala Gly Thr Pro Ile
                290                 295                 300

GGA TCA CCC ATG ATC AAT TCC AAA ACC AAC AAT AAT AAT AAT GAC CCT    1382
Gly Ser Pro Met Ile Asn Ser Lys Thr Asn Asn Asn Asn Asn Asp Pro
            305                 310                 315

AAC AAT TAC TCA TCA ACC AAA AAC AAT GTC CAA GAG GCA AAT TTA CAA    1430
Asn Asn Tyr Ser Ser Thr Lys Asn Asn Val Gln Glu Ala Asn Leu Gln
        320                 325                 330

GAA TGG GTA AAA CCT CCA GCA AAA TCT ACT GTC TCA CAA TTC AAA CCT    1478
Glu Trp Val Lys Pro Pro Ala Lys Ser Thr Val Ser Gln Phe Lys Pro
    335                 340                 345

AGT CGA GCT GCA CCA AAA CCA CCA ACT CCA TAT CAT TTG ACA CAA CTA    1526
Ser Arg Ala Ala Pro Lys Pro Pro Thr Pro Tyr His Leu Thr Gln Leu
350                 355                 360                 365

AAT GGC TCT TCC CAC CAA CAT ACA TCA TCA TCA GGC TCA TTA CCT AGT    1574
Asn Gly Ser Ser His Gln His Thr Ser Ser Ser Gly Ser Leu Pro Ser
                370                 375                 380

TCT GGT AAT AAT AAT AAT AAT AAC AGC ACT AAC AAT AAT AAT ACT AAA    1622
Ser Gly Asn Asn Asn Asn Asn Asn Ser Thr Asn Asn Asn Asn Thr Lys
            385                 390                 395

AAC GTT TCA CCA TTG AAT AAT TTG ATG AAT AAA TCT GAA CTT ATT CCT    1670
Asn Val Ser Pro Leu Asn Asn Leu Met Asn Lys Ser Glu Leu Ile Pro
        400                 405                 410

GCT AGA AGA GCT CCA CCA CCT CCA ACA AGT GGC ACA TCT TCA GAT ACA    1718
Ala Arg Arg Ala Pro Pro Pro Pro Thr Ser Gly Thr Ser Ser Asp Thr
    415                 420                 425

TAT TCT AAT AAG AAT CAT CAA GAT AGA TCT GGA TAT GAA CAA CAA CGT    1766
Tyr Ser Asn Lys Asn His Gln Asp Arg Ser Gly Tyr Glu Gln Gln Arg
430                 435                 440                 445

CAA CAA CGT ACT GAC TCA TCA CAA CAA CAA CAA CAA CAA AAG CAA CAT    1814
Gln Gln Arg Thr Asp Ser Ser Gln Gln Gln Gln Gln Gln Lys Gln His
                450                 455                 460
```

```
CAA TAT CAA CAG AAA TCC CAA CAA CAA CAA CAA CAA CCA CAA CAA CCA    1862
Gln Tyr Gln Gln Lys Ser Gln Gln Gln Gln Gln Pro Gln Gln Pro
            465                 470                 475

TTA TCT CTG CAT CAA GGT GGG ACT TCG CAT ATT CCG AAA CAA GTA CCT    1910
Leu Ser Leu His Gln Gly Gly Thr Ser His Ile Pro Lys Gln Val Pro
        480                 485                 490

CCT ACA TTA CCA TCA TCT GGA CCA CCC ACT CAG GCT GCT TCA GGA AAA    1958
Pro Thr Leu Pro Ser Ser Gly Pro Pro Thr Gln Ala Ala Ser Gly Lys
    495                 500                 505

TCA ATG CCA TCT AAA ATT CAT CCT GAT CTT AAG ATT CAA CAA GGC ACA    2006
Ser Met Pro Ser Lys Ile His Pro Asp Leu Lys Ile Gln Gln Gly Thr
510                 515                 520                 525

AAT AAT TAT ATT AAG AGT AGC GGT ACT GAT GCT AAT CAA GTC GAT GGT    2054
Asn Asn Tyr Ile Lys Ser Ser Gly Thr Asp Ala Asn Gln Val Asp Gly
                530                 535                 540

GAT GCT AAA CAA TTT ATT AAA CCA TTT AAT TTA CAA CTG AAA AAG AGT    2102
Asp Ala Lys Gln Phe Ile Lys Pro Phe Asn Leu Gln Leu Lys Lys Ser
            545                 550                 555

CAG CAA CAA TTG GCA TCA AAA CAA CCG TCA CCA CCT TCA TCT CAA CAA    2150
Gln Gln Gln Leu Ala Ser Lys Gln Pro Ser Pro Pro Ser Ser Gln Gln
        560                 565                 570

CAG CAA CAA AAA CCT ATG ACA TCA CAT GGA TTA ATG GGT ACA TCA CAT    2198
Gln Gln Gln Lys Pro Met Thr Ser His Gly Leu Met Gly Thr Ser His
    575                 580                 585

TCA GTT ACT AAA CCA TTG AAT CCA GTC AAT GAT CCA ATC AAA CCA TTA    2246
Ser Val Thr Lys Pro Leu Asn Pro Val Asn Asp Pro Ile Lys Pro Leu
590                 595                 600                 605

AAT TTG AAA TCA TCT AAA TCC AAA GAA GCA TTA AAT GAA ACT CTG GGG    2294
Asn Leu Lys Ser Ser Lys Ser Lys Glu Ala Leu Asn Glu Thr Leu Gly
                610                 615                 620

GTG CTG AAA ACA CCA TCA CCT ACA GAT AAA TCA AAT AAA CCA ACT GCA    2342
Val Leu Lys Thr Pro Ser Pro Thr Asp Lys Ser Asn Lys Pro Thr Ala
            625                 630                 635

CCT GCT AGT GGT CCT GCA GTG ACC AAA ACA GCT AAA CAA CTC AAG AAG    2390
Pro Ala Ser Gly Pro Ala Val Thr Lys Thr Ala Lys Gln Leu Lys Lys
        640                 645                 650

GAA CGA GAA AGA TTG AAT GAT TTA CAA ATC ATT GCT AAA TTG AAA ACA    2438
Glu Arg Glu Arg Leu Asn Asp Leu Gln Ile Ile Ala Lys Leu Lys Thr
    655                 660                 665

GTG GTT AAT AAT CAA GAT CCT AAA CCA TTA TTT AGA ATT GTT GAA AAA    2486
Val Val Asn Asn Gln Asp Pro Lys Pro Leu Phe Arg Ile Val Glu Lys
670                 675                 680                 685
```

```
GCT GGT CAA GGT GCT AGT GGG AAT GTT TAT TTG GCG GAA ATG ATC AAA     2534
Ala Gly Gln Gly Ala Ser Gly Asn Val Tyr Leu Ala Glu Met Ile Lys
                690                 695                 700

GAT AAT AAT CGA AAG ATT GCG ATT AAA CAA ATG GAT CTT GAT GCT CAA     2582
Asp Asn Asn Arg Lys Ile Ala Ile Lys Gln Met Asp Leu Asp Ala Gln
            705                 710                 715

CCC CGT AAA GAG TTA ATA ATA AAT GAA ATC TTG GTT ATG AAA GAT AGT     2630
Pro Arg Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Lys Asp Ser
        720                 725                 730

CAA CAT AAA AAT ATT GTT AAT TTT TTG GAT TCT TAT TTA ATT GGT GAT     2678
Gln His Lys Asn Ile Val Asn Phe Leu Asp Ser Tyr Leu Ile Gly Asp
    735                 740                 745

AAT GAA TTA TGG GTA ATT ATG GAA TAT ATG CAA GGT GGT TCA TTA ACG     2726
Asn Glu Leu Trp Val Ile Met Glu Tyr Met Gln Gly Gly Ser Leu Thr
750                 755                 760                 765

GAA ATC ATT GAA AAT AAT GAT TTT AAA TTG AAT GAA AAA CAA ATT GCC     2774
Glu Ile Ile Glu Asn Asn Asp Phe Lys Leu Asn Glu Lys Gln Ile Ala
                770                 775                 780

ACT ATA TGT TTT GAA ACC TTA AAG GGG TTA CAA CAT TTA CAT AAA AAA     2822
Thr Ile Cys Phe Glu Thr Leu Lys Gly Leu Gln His Leu His Lys Lys
            785                 790                 795

CAT ATT ATT CAT CGT GAT ATT AAA TCC GAT AAT GTT TTA TTA GAT GCA     2870
His Ile Ile His Arg Asp Ile Lys Ser Asp Asn Val Leu Leu Asp Ala
        800                 805                 810

TAT GGT AAT GTT AAA ATC ACT GAT TTT GGA TTT TGT GCT AAA TTA ACT     2918
Tyr Gly Asn Val Lys Ile Thr Asp Phe Gly Phe Cys Ala Lys Leu Thr
    815                 820                 825

GAT CAA AGA AAT AAA CGT GCC ACA ATG GTG GGG ACA CCA TAT TGG ATG     2966
Asp Gln Arg Asn Lys Arg Ala Thr Met Val Gly Thr Pro Tyr Trp Met
830                 835                 840                 845

GCA CCT GAA GTG GTT AAA CAA AAG GAA TAT GAT GAA AAA GTT GAT GTT     3014
Ala Pro Glu Val Val Lys Gln Lys Glu Tyr Asp Glu Lys Val Asp Val
                850                 855                 860

TGG TCA TTG GGG ATT ATG ACT ATT GAA ATG ATT GAA GGA GAA CCA CCT     3062
Trp Ser Leu Gly Ile Met Thr Ile Glu Met Ile Glu Gly Glu Pro Pro
            865                 870                 875

TAT TTG AAT GAA GAA CCA TTA AAA GCT TTA TAT CTT ATA GCT ACT AAT     3110
Tyr Leu Asn Glu Glu Pro Leu Lys Ala Leu Tyr Leu Ile Ala Thr Asn
        880                 885                 890

GGT ACA CCA AAA TTG AAA AAA CCC GAA TTA TTA TCG AAT TCA ATT AAA     3158
Gly Thr Pro Lys Leu Lys Lys Pro Glu Leu Leu Ser Asn Ser Ile Lys
    895                 900                 905

AAA TTC TTA TCA ATT TGT CTT TGT GTT GAT GTT AGA TAT CGT GCT AGT     3206
Lys Phe Leu Ser Ile Cys Leu Cys Val Asp Val Arg Tyr Arg Ala Ser
910                 915                 920                 925
```

```
ACT GAT GAA TTA TTA GAA CAT TCA TTT ATT CAA CAT AAA TCA GGG AAA    3254
Thr Asp Glu Leu Leu Glu His Ser Phe Ile Gln His Lys Ser Gly Lys
                930                 935                 940

ATT GAA GAA TTG GCA CCA TTA TTA GAA TGG AAA AAA CAA CAA CAA AAG    3302
Ile Glu Glu Leu Ala Pro Leu Leu Glu Trp Lys Lys Gln Gln Gln Lys
            945                 950                 955

CAT CAA CAG CAT AAA CAA GAA ACA CTG GAT ACA GGA TTT GCA TAGAGATTG  3353
His Gln Gln His Lys Gln Glu Thr Leu Asp Thr Gly Phe Ala
        960                 965                 970

AATATAGCCG TAGAAAACTG GTACTTTGGT TTTGGTATAA TATATTTATG TGATGTGTTG   3413
TGTGTATGGT TAGTTTAGAT TTGGATTTTT AGTTTTTTAG AGTTTAGTTT TTCAATTTTT   3473
AGTTTTAGAG ACAATATTCT AGA                                           3496
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 971 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Thr Ser Ile Tyr Thr Ser Asp Leu Lys Asn His Arg Arg Ala Pro
 1               5                  10                  15
Pro Pro Pro Asn Gly Ala Ala Gly Ser Gly Ser Gly Ser Gly Ser Gly
            20                  25                  30
Ser Gly Ser Gly Ser Gly Ser Leu Ala Asn Ile Val Thr Ser Ser Asn
        35                  40                  45
Ser Leu Gly Val Thr Ala Asn Gln Thr Lys Pro Ile Gln Leu Asn Ile
    50                  55                  60
Asn Ser Ser Lys Arg Gln Ser Gly Trp Val His Val Lys Asp Asp Gly
65                  70                  75                  80
Ile Phe Thr Ser Phe Arg Trp Asn Lys Arg Phe Met Val Ile Asn Asp
                85                  90                  95
Lys Thr Leu Asn Phe Tyr Lys Gln Glu Pro Tyr Ser Ser Asp Gly Asn
            100                 105                 110
Ser Asn Ser Asn Thr Pro Asp Leu Ser Phe Pro Leu Tyr Leu Ile Asn
        115                 120                 125
Asn Ile Asn Leu Lys Pro Asn Ser Gly Tyr Ser Lys Thr Ser Gln Ser
    130                 135                 140
Phe Glu Ile Val Pro Lys Asn Asn Asn Lys Ser Ile Leu Ile Ser Val
145                 150                 155                 160
Lys Thr Asn Asn Asp Tyr Leu Asp Trp Leu Asp Ala Phe Thr Thr Lys
                165                 170                 175
Cys Pro Leu Val Gln Ile Gly Glu Asn Asn Ser Gly Val Ser Ser Ser
            180                 185                 190
His Pro His Leu Gln Ile Gln His Leu Thr Asn Gly Ser Leu Asn Gly
        195                 200                 205
Asn Ser Ser Ser Ser Pro Thr Ser Gly Leu Leu Ser Ser Ser Val Leu
    210                 215                 220
```

```
Thr Gly Gly Asn Ser Gly Val Ser Gly Pro Ile Asn Phe Thr His Lys
225                 230                 235                 240
Val His Val Gly Phe Asp Pro Ala Ser Gly Asn Phe Thr Gly Leu Pro
                245                 250                 255
Asp Thr Trp Lys Ser Leu Leu Gln His Ser Lys Ile Thr Asn Glu Asp
            260                 265                 270
Trp Lys Lys Asp Pro Val Ala Val Ile Glu Val Leu Glu Phe Tyr Ser
        275                 280                 285
Asp Ile Asn Gly Gly Asn Ser Ala Ala Gly Thr Pro Ile Gly Ser Pro
    290                 295                 300
Met Ile Asn Ser Lys Thr Asn Asn Asn Asn Asn Asp Pro Asn Asn Tyr
305                 310                 315                 320
Ser Ser Thr Lys Asn Asn Val Gln Glu Ala Asn Leu Gln Glu Trp Val
                325                 330                 335
Lys Pro Pro Ala Lys Ser Thr Val Ser Gln Phe Lys Pro Ser Arg Ala
            340                 345                 350
Ala Pro Lys Pro Pro Thr Pro Tyr His Leu Thr Gln Leu Asn Gly Ser
        355                 360                 365
Ser His Gln His Thr Ser Ser Ser Gly Ser Leu Pro Ser Ser Gly Asn
    370                 375                 380
Asn Asn Asn Asn Asn Ser Thr Asn Asn Asn Asn Thr Lys Asn Val Ser
385                 390                 395                 400
Pro Leu Asn Asn Leu Met Asn Lys Ser Glu Leu Ile Pro Ala Arg Arg
                405                 410                 415
Ala Pro Pro Pro Pro Thr Ser Gly Thr Ser Ser Asp Thr Tyr Ser Asn
            420                 425                 430
Lys Asn His Gln Asp Arg Ser Gly Tyr Glu Gln Gln Arg Gln Gln Arg
        435                 440                 445
Thr Asp Ser Ser Gln Gln Gln Gln Gln Gln Lys Gln His Gln Tyr Gln
    450                 455                 460
Gln Lys Ser Gln Gln Gln Gln Gln Gln Pro Gln Gln Pro Leu Ser Leu
465                 470                 475                 480
His Gln Gly Gly Thr Ser His Ile Pro Lys Gln Val Pro Pro Thr Leu
                485                 490                 495
Pro Ser Ser Gly Pro Pro Thr Gln Ala Ala Ser Gly Lys Ser Met Pro
            500                 505                 510
Ser Lys Ile His Pro Asp Leu Lys Ile Gln Gln Gly Thr Asn Asn Tyr
        515                 520                 525
Ile Lys Ser Ser Gly Thr Asp Ala Asn Gln Val Asp Gly Asp Ala Lys
    530                 535                 540
Gln Phe Ile Lys Pro Phe Asn Leu Gln Leu Lys Lys Ser Gln Gln Gln
545                 550                 555                 560
Leu Ala Ser Lys Gln Pro Ser Pro Pro Ser Ser Gln Gln Gln Gln Gln
                565                 570                 575
Lys Pro Met Thr Ser His Gly Leu Met Gly Thr Ser His Ser Val Thr
            580                 585                 590
Lys Pro Leu Asn Pro Val Asn Asp Pro Ile Lys Pro Leu Asn Leu Lys
        595                 600                 605
Ser Ser Lys Ser Lys Glu Ala Leu Asn Glu Thr Leu Gly Val Leu Lys
    610                 615                 620
Thr Pro Ser Pro Thr Asp Lys Ser Asn Lys Pro Thr Ala Pro Ala Ser
625                 630                 635                 640
Gly Pro Ala Val Thr Lys Thr Ala Lys Gln Leu Lys Lys Glu Arg Glu
                645                 650                 655
Arg Leu Asn Asp Leu Gln Ile Ile Ala Lys Leu Lys Thr Val Val Asn
            660                 665                 670
Asn Gln Asp Pro Lys Pro Leu Phe Arg Ile Val Glu Lys Ala Gly Gln
        675                 680                 685
```

```
Gly Ala Ser Gly Asn Val Tyr Leu Ala Glu Met Ile Lys Asp Asn Asn
    690                 695                 700
Arg Lys Ile Ala Ile Lys Gln Met Asp Leu Asp Ala Gln Pro Arg Lys
705                 710                 715                 720
Glu Leu Ile Ile Asn Glu Ile Leu Val Met Lys Asp Ser Gln His Lys
                725                 730                 735
Asn Ile Val Asn Phe Leu Asp Ser Tyr Leu Ile Gly Asp Asn Glu Leu
            740                 745                 750
Trp Val Ile Met Glu Tyr Met Gln Gly Gly Ser Leu Thr Glu Ile Ile
        755                 760                 765
Glu Asn Asn Asp Phe Lys Leu Asn Glu Lys Gln Ile Ala Thr Ile Cys
    770                 775                 780
Phe Glu Thr Leu Lys Gly Leu Gln His Leu His Lys Lys His Ile Ile
785                 790                 795                 800
His Arg Asp Ile Lys Ser Asp Asn Val Leu Leu Asp Ala Tyr Gly Asn
                805                 810                 815
Val Lys Ile Thr Asp Phe Gly Phe Cys Ala Lys Leu Thr Asp Gln Arg
            820                 825                 830
Asn Lys Arg Ala Thr Met Val Gly Thr Pro Tyr Trp Met Ala Pro Glu
        835                 840                 845
Val Val Lys Gln Lys Glu Tyr Asp Glu Lys Val Asp Val Trp Ser Leu
    850                 855                 860
Gly Ile Met Thr Ile Glu Met Ile Glu Gly Glu Pro Pro Tyr Leu Asn
865                 870                 875                 880
Glu Glu Pro Leu Lys Ala Leu Tyr Leu Ile Ala Thr Asn Gly Thr Pro
                885                 890                 895
Lys Leu Lys Lys Pro Glu Leu Leu Ser Asn Ser Ile Lys Lys Phe Leu
            900                 905                 910
Ser Ile Cys Leu Cys Val Asp Val Arg Tyr Arg Ala Ser Thr Asp Glu
        915                 920                 925
Leu Leu Glu His Ser Phe Ile Gln His Lys Ser Gly Lys Ile Glu Glu
    930                 935                 940
Leu Ala Pro Leu Leu Glu Trp Lys Lys Gln Gln Gln Lys His Gln Gln
945                 950                 955                 960
His Lys Gln Glu Thr Leu Asp Thr Gly Phe Ala
                965                 970
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1031 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 271...843
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
CAACCAAACC AACTTTCATC CTTCTACCAA TATCTTCAAC AAAAGTTTTA TTCAATACTA     60
TTTTAAAAAT AACAGTGTTA CTCGTTCATT TGATTTGTTA ATAAGACTGA TTTACCCACT    120
TTTTAGTTCC TATAATCATA CAGATTTCTC GTCCTAAATC TATTTTTATT GTTATTTTTA    180
CTTTAGTTTT CACTTTTGCT TTCAGTTTTT TCTTTTTTTA GCACAAGAGA AAAGTATTCA    240
```

```
GCTCATAAAT AATTAATATA TCCATATATC  ATG CAA ACT ATA AAA TGT GTT GTT    294
                                  Met Gln Thr Ile Lys Cys Val Val
                                   1               5

GTC GGT GAT GGT GCC GTT GGT AAA ACT TGC TTA TTA ATC TCG TAT ACC    342
Val Gly Asp Gly Ala Val Gly Lys Thr Cys Leu Leu Ile Ser Tyr Thr
    10                  15                  20

ACT AGT AAA TTT CCA GCT GAT TAT GTT CCT ACT GTT TTT GAT AAT TAT    390
Thr Ser Lys Phe Pro Ala Asp Tyr Val Pro Thr Val Phe Asp Asn Tyr
25                  30                  35                  40

GCT GTA ACC GTG ATG ATA GGA GAC GAA CCA TTT ACC TTG GGA TTA TTT    438
Ala Val Thr Val Met Ile Gly Asp Glu Pro Phe Thr Leu Gly Leu Phe
                45                  50                  55

GAT ACT GCT GGT CAA GAA GAT TAC GAC AGA TTA AGG CCT TTG TCA TAT    486
Asp Thr Ala Gly Gln Glu Asp Tyr Asp Arg Leu Arg Pro Leu Ser Tyr
            60                  65                  70

CCA TCG ACT GAT GTA TTC CTT GTT TGT TTT TCC GTC ATT TCT CCC GCT    534
Pro Ser Thr Asp Val Phe Leu Val Cys Phe Ser Val Ile Ser Pro Ala
        75                  80                  85

TCG TTT GAA AAT GTT AAA GAA AAA TGG TTC CCA GAA GTT CAT CAC CAT    582
Ser Phe Glu Asn Val Lys Glu Lys Trp Phe Pro Glu Val His His His
    90                  95                  100

TGT CCC GGT GTG CCA ATA ATT ATT GTC GGT ACC CAA ACT GAT TTA CGA    630
Cys Pro Gly Val Pro Ile Ile Ile Val Gly Thr Gln Thr Asp Leu Arg
105                 110                 115                 120

AAC GAT GAT GTT ATT TTA CAG AGA TTG CAC AGA CAA AAA TTG TCC CCA    678
Asn Asp Asp Val Ile Leu Gln Arg Leu His Arg Gln Lys Leu Ser Pro
                125                 130                 135

ATC ACC CAG GAA CAG GGT GAA AAA TTG GCT AAG GAA TTG AGA GCT GTC    726
Ile Thr Gln Glu Gln Gly Glu Lys Leu Ala Lys Glu Leu Arg Ala Val
            140                 145                 150

AAG TAT GTT GAG TGT TCT GCA TTG ACT CAA AGA GGA TTG AAA ACA GTG    774
Lys Tyr Val Glu Cys Ser Ala Leu Thr Gln Arg Gly Leu Lys Thr Val
        155                 160                 165

TTT GAC GAG GCT ATA GTA GCT GCA TTA GAA CCT CCT GTA ATT AAA AAA    822
Phe Asp Glu Ala Ile Val Ala Ala Leu Glu Pro Pro Val Ile Lys Lys
    170                 175                 180

TCG AAA AAG TGT ACT ATT TTA TAGGTCGGCG ATACTAGAAG ATAGAGGATA TTGG  877
Ser Lys Lys Cys Thr Ile Leu
185                 190

AAATAGGGCA TACATGAGAT ATTGAATATC TATCATTAAA TATATAATTA GTTTTTTTCT   937
AAAACCTATC TTTAGGTTTG ATCTCGTTTG ATGTGTTGGG CTGTTTCGCA AAACAGTGTT   997
CCAATCAATA AAAAGATGTG TGTAAGACTC TAGA                              1031
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Gln Thr Ile Lys Cys Val Val Val Gly Asp Gly Ala Val Gly Lys
 1               5                  10                  15
Thr Cys Leu Leu Ile Ser Tyr Thr Thr Ser Lys Phe Pro Ala Asp Tyr
            20                  25                  30
Val Pro Thr Val Phe Asp Asn Tyr Ala Val Thr Val Met Ile Gly Asp
        35                  40                  45
Glu Pro Phe Thr Leu Gly Leu Phe Asp Thr Ala Gly Gln Glu Asp Tyr
    50                  55                  60
Asp Arg Leu Arg Pro Leu Ser Tyr Pro Ser Thr Asp Val Phe Leu Val
65                  70                  75                  80
Cys Phe Ser Val Ile Ser Pro Ala Ser Phe Glu Asn Val Lys Glu Lys
                85                  90                  95
Trp Phe Pro Glu Val His His His Cys Pro Gly Val Pro Ile Ile Ile
            100                 105                 110
Val Gly Thr Gln Thr Asp Leu Arg Asn Asp Asp Val Ile Leu Gln Arg
        115                 120                 125
Leu His Arg Gln Lys Leu Ser Pro Ile Thr Gln Glu Gln Gly Glu Lys
    130                 135                 140
Leu Ala Lys Glu Leu Arg Ala Val Lys Tyr Val Glu Cys Ser Ala Leu
145                 150                 155                 160
Thr Gln Arg Gly Leu Lys Thr Val Phe Asp Glu Ala Ile Val Ala Ala
                165                 170                 175
Leu Glu Pro Pro Val Ile Lys Lys Ser Lys Lys Cys Thr Ile Leu
            180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2231 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 291...2195
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
AAGCTTGTTT CTTATCTCCT TAGTATATTG TTTTACAACA CCACATACAC ATACACATAT     60
AGCCTTCATT AGCCTTCATT TTGACATATT TCAATAACAA TCAAGAACTA CAAGTCATAA    120
CTGACACACA TATAATATCT TAATTGTTAT TATAAATTTA TTCTTGATTA GATTTTAGAC    180
```

```
GGGCAGAAAC AAAAACGGAA AATCCAACTC ATCCCCGATA ACTACACACA TCTATATTAA     240
ATCATCTATT AGTCTATCAG TTATATCTCC CTCCCCTTTT CTTCTAACAA  ATG ATT      296
                                                        Met Ile
                                                         1

AAG ACG TTT CGG AAA AGT AAA AGA CTG TCG AGT AAT TCA AGT TCA CCC      344
Lys Thr Phe Arg Lys Ser Lys Arg Leu Ser Ser Asn Ser Ser Ser Pro
          5                  10                  15

AAG AAA ACA ATA TCT CGA GTA TCA TCA ACT TCA AGT AAT CAA ACA TCT      392
Lys Lys Thr Ile Ser Arg Val Ser Ser Thr Ser Ser Asn Gln Thr Ser
     20                  25                  30

CAT GAT GGA ATA TTA CAA TCA CCT AAA AAA GTC ATT AGA GCT CTA TAT      440
His Asp Gly Ile Leu Gln Ser Pro Lys Lys Val Ile Arg Ala Leu Tyr
35                  40                  45                  50

GAT TAT GAA CCT CAA GGT CCT GGA GAA TTG AAA TTT TTC AAA GGA GAT      488
Asp Tyr Glu Pro Gln Gly Pro Gly Glu Leu Lys Phe Phe Lys Gly Asp
                55                  60                  65

TTT TTC CAT GTA TTA AAT GAT GTT GAT GAT GAA TTA CAT AAA GAA GCG      536
Phe Phe His Val Leu Asn Asp Val Asp Asp Glu Leu His Lys Glu Ala
           70                  75                  80

GAA CGT AAT GGA TGG ATA GAA GCA ACA AAT CCA ATG ACT CAA CTT AAA      584
Glu Arg Asn Gly Trp Ile Glu Ala Thr Asn Pro Met Thr Gln Leu Lys
      85                  90                  95

GGG ATG GTC CCC ATT AGT TAT TTT GAA ATA TTT GAT CGA TCT CGT CCT      632
Gly Met Val Pro Ile Ser Tyr Phe Glu Ile Phe Asp Arg Ser Arg Pro
     100                 105                 110

ACA GTT ACA GCA TCA TCA AAC AGT TTT ACA AAT TCC ATT GAT ATT CAA      680
Thr Val Thr Ala Ser Ser Asn Ser Phe Thr Asn Ser Ile Asp Ile Gln
115                 120                 125                 130

CAT CAA CAT CAA CAA GGA ATT CAC AAT GGA ACA GGA AAT CGA AAT TTA      728
His Gln His Gln Gln Gly Ile His Asn Gly Thr Gly Asn Arg Asn Leu
                135                 140                 145

AAT CAA ACA TTA TAT GCT GTT ACA CTA TAT GAA TTT AAA GCT GAA CGA      776
Asn Gln Thr Leu Tyr Ala Val Thr Leu Tyr Glu Phe Lys Ala Glu Arg
           150                 155                 160

GAT GAT GAA TTG GAT ATA ATG CCT AAT GAA AAT TTA ATT ATT TGT GCA      824
Asp Asp Glu Leu Asp Ile Met Pro Asn Glu Asn Leu Ile Ile Cys Ala
      165                 170                 175

CAT CAT GAT TAT GAA TGG TTT ATT GCC AAA CCA ATA AAT CGA TTA GGT      872
His His Asp Tyr Glu Trp Phe Ile Ala Lys Pro Ile Asn Arg Leu Gly
     180                 185                 190

GGA CCA GGT TTA GTA CCT GTT TCT TAT GTT AAA ATA ATT GAT CTT TTA      920
Gly Pro Gly Leu Val Pro Val Ser Tyr Val Lys Ile Ile Asp Leu Leu
195                 200                 205                 210
```

```
AAC CCT AAT TCT CAT TAT ACA TCA ATT GAT ACA TCA AGG CGA TCA CAA      968
Asn Pro Asn Ser His Tyr Thr Ser Ile Asp Thr Ser Arg Arg Ser Gln
                215                 220                 225

GTC ATA CAA GTA ATC AAT GGA TTT AAT ATA CCG ACA GTA GAA CAA TGG     1016
Val Ile Gln Val Ile Asn Gly Phe Asn Ile Pro Thr Val Glu Gln Trp
            230                 235                 240

AAA AAT CAA ACT GCC AAA TAT CAA GCT TCA ACA ATC CCC CTT GGT TCA     1064
Lys Asn Gln Thr Ala Lys Tyr Gln Ala Ser Thr Ile Pro Leu Gly Ser
        245                 250                 255

ATA TCA GGA AGT GGT ACT CCA CCA ACA TCA GCT AAT TCA CAA TAT TTT     1112
Ile Ser Gly Ser Gly Thr Pro Pro Thr Ser Ala Asn Ser Gln Tyr Phe
    260                 265                 270

GAT AAT CAT ACT ATG ACT TCA AAT CGA TCA TCA CTG GGT TCA TCA ATT     1160
Asp Asn His Thr Met Thr Ser Asn Arg Ser Ser Leu Gly Ser Ser Ile
275                 280                 285                 290

TCT ATT ATT GAA GCT AGT GTT GAT TCA TAT CAA TTA GAT CAT GGT CGA     1208
Ser Ile Ile Glu Ala Ser Val Asp Ser Tyr Gln Leu Asp His Gly Arg
                295                 300                 305

TAT CAA TAT TCA ATA ACT GCT CGA TTA AAT AAT GGC AGA ATA AGA TAT     1256
Tyr Gln Tyr Ser Ile Thr Ala Arg Leu Asn Asn Gly Arg Ile Arg Tyr
            310                 315                 320

TTA TAT CGA TAT TAT CAA GAT TTT TAT GAT TTA CAA GTG AAA TTA TTA     1304
Leu Tyr Arg Tyr Tyr Gln Asp Phe Tyr Asp Leu Gln Val Lys Leu Leu
        325                 330                 335

GAA TTA TTT CCT TAT GAA GCT GGG AGA ATT GAA AAT TCT AAA AGA ATA     1352
Glu Leu Phe Pro Tyr Glu Ala Gly Arg Ile Glu Asn Ser Lys Arg Ile
    340                 345                 350

ATT CCA TCT ATA CCA GGA CCT TTA ATT AAT GTC AAT GAT TCA ATA TCA     1400
Ile Pro Ser Ile Pro Gly Pro Leu Ile Asn Val Asn Asp Ser Ile Ser
355                 360                 365                 370

AAA TTA CGA AGA GAA AAA TTG GAT TAT TAT TTA TCA AAT TTA ATT GCA     1448
Lys Leu Arg Arg Glu Lys Leu Asp Tyr Tyr Leu Ser Asn Leu Ile Ala
                375                 380                 385

TTA CCT AGT CAT ATA TCT CGA TCA GAA GAA GTA TTA AAA TTA TTT GAT     1496
Leu Pro Ser His Ile Ser Arg Ser Glu Glu Val Leu Lys Leu Phe Asp
            390                 395                 400

GTT TTA GAT AAT GGA TTT GAT CGA GAA ACT GAT GCT ATT AAT AAA CGA     1544
Val Leu Asp Asn Gly Phe Asp Arg Glu Thr Asp Ala Ile Asn Lys Arg
        405                 410                 415

TTT TCT AAA CCA ATA AGT CAA AAA TCA AAT TCT CAT CAA GAT AGA TTA     1592
Phe Ser Lys Pro Ile Ser Gln Lys Ser Asn Ser His Gln Asp Arg Leu
    420                 425                 430

TCT CAA TAT TCC AAT TTT AAC GTT TTA CAA CAA CAA CAA CAA CAA CAG     1640
Ser Gln Tyr Ser Asn Phe Asn Val Leu Gln Gln Gln Gln Gln Gln Gln
435                 440                 445                 450
```

```
CAA CAA CAG CAA TAT GCT CAT CAT TCA AGA GGT TCT GAT AAT TCA CCT      1688
Gln Gln Gln Gln Tyr Ala His His Ser Arg Gly Ser Asp Asn Ser Pro
                455                 460                 465

ACT AAT GAA TCA TCA GGT TCA AAT TTA ATT AAT TCT TCT TCT CAT AAT      1736
Thr Asn Glu Ser Ser Gly Ser Asn Leu Ile Asn Ser Ser Ser His Asn
            470                 475                 480

GAT TCA TCA TTA TCT TCA TCA CCA CCA CCA CCA CCA CCA CAA ACT GTC      1784
Asp Ser Ser Leu Ser Ser Ser Pro Pro Pro Pro Pro Pro Gln Thr Val
        485                 490                 495

ACC ACC ACG AAC ACC ACG AAC ACC ACC ATA ACC ACA GAC TCC TCA TCA      1832
Thr Thr Thr Asn Thr Thr Asn Thr Thr Ile Thr Thr Asp Ser Ser Ser
    500                 505                 510

AAA CAA CCA AAA GCC AAA GTG AAA TTT TAT TTT GAT GAT GAT ATA TTT      1880
Lys Gln Pro Lys Ala Lys Val Lys Phe Tyr Phe Asp Asp Asp Ile Phe
515                 520                 525                 530

GTA TTA TTA ATC CCA ACC AAT TTA CGA TTA CAA GAT TTA AAA TCA AAA      1928
Val Leu Leu Ile Pro Thr Asn Leu Arg Leu Gln Asp Leu Lys Ser Lys
                535                 540                 545

TTA TTT AAA CGA TTA GAA TTG GAT ATT ACT TAT AAA TAT GAA AAA CCT      1976
Leu Phe Lys Arg Leu Glu Leu Asp Ile Thr Tyr Lys Tyr Glu Lys Pro
            550                 555                 560

GAT CAA CAA CAA AAA CCT ACA TCA GAA TCA ATT CAT TTA TTT TTG AAA      2024
Asp Gln Gln Gln Lys Pro Thr Ser Glu Ser Ile His Leu Phe Leu Lys
        565                 570                 575

AAT GAT TTT GAA GAT TTT TTA ATT GAA AAT GAA ACT AGC AAC AAC AAC      2072
Asn Asp Phe Glu Asp Phe Leu Ile Glu Asn Glu Thr Ser Asn Asn Asn
    580                 585                 590

AAT CTG GAA ATT GAT TTC GAA AAT GAA ATT ATT AAA GAA AAA TTA GGA      2120
Asn Leu Glu Ile Asp Phe Glu Asn Glu Ile Ile Lys Glu Lys Leu Gly
595                 600                 605                 610

GAA TTT GAA GTT AAT GAT GAT GAA AAA TTT CAA AGT ATT TTA TTT GAT      2168
Glu Phe Glu Val Asn Asp Asp Glu Lys Phe Gln Ser Ile Leu Phe Asp
                615                 620                 625

AAA TGT AAA TTA ATG GTT TTA GTA TAT TAAACAGAGA TCAATAAGAG AGAGAGA    2222
Lys Cys Lys Leu Met Val Leu Val Tyr
            630                 635

GAGAGACAT                                                           2231
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 635 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Ile Lys Thr Phe Arg Lys Ser Lys Arg Leu Ser Ser Asn Ser Ser
 1               5                  10                  15
Ser Pro Lys Lys Thr Ile Ser Arg Val Ser Ser Thr Ser Ser Asn Gln
            20                  25                  30
Thr Ser His Asp Gly Ile Leu Gln Ser Pro Lys Lys Val Ile Arg Ala
        35                  40                  45
Leu Tyr Asp Tyr Glu Pro Gln Gly Pro Gly Glu Leu Lys Phe Phe Lys
    50                  55                  60
Gly Asp Phe Phe His Val Leu Asn Asp Val Asp Asp Glu Leu His Lys
65                  70                  75                  80
Glu Ala Glu Arg Asn Gly Trp Ile Glu Ala Thr Asn Pro Met Thr Gln
                85                  90                  95
Leu Lys Gly Met Val Pro Ile Ser Tyr Phe Glu Ile Phe Asp Arg Ser
            100                 105                 110
Arg Pro Thr Val Thr Ala Ser Ser Asn Ser Phe Thr Asn Ser Ile Asp
        115                 120                 125
Ile Gln His Gln His Gln Gln Gly Ile His Asn Gly Thr Gly Asn Arg
    130                 135                 140
Asn Leu Asn Gln Thr Leu Tyr Ala Val Thr Leu Tyr Glu Phe Lys Ala
145                 150                 155                 160
Glu Arg Asp Asp Glu Leu Asp Ile Met Pro Asn Glu Asn Leu Ile Ile
                165                 170                 175
Cys Ala His His Asp Tyr Glu Trp Phe Ile Ala Lys Pro Ile Asn Arg
            180                 185                 190
Leu Gly Gly Pro Gly Leu Val Pro Val Ser Tyr Val Lys Ile Ile Asp
        195                 200                 205
Leu Leu Asn Pro Asn Ser His Tyr Thr Ser Ile Asp Thr Ser Arg Arg
    210                 215                 220
Ser Gln Val Ile Gln Val Ile Asn Gly Phe Asn Ile Pro Thr Val Glu
225                 230                 235                 240
Gln Trp Lys Asn Gln Thr Ala Lys Tyr Gln Ala Ser Thr Ile Pro Leu
                245                 250                 255
Gly Ser Ile Ser Gly Ser Gly Thr Pro Pro Thr Ser Ala Asn Ser Gln
            260                 265                 270
Tyr Phe Asp Asn His Thr Met Thr Ser Asn Arg Ser Ser Leu Gly Ser
        275                 280                 285
Ser Ile Ser Ile Ile Glu Ala Ser Val Asp Ser Tyr Gln Leu Asp His
    290                 295                 300
Gly Arg Tyr Gln Tyr Ser Ile Thr Ala Arg Leu Asn Asn Gly Arg Ile
305                 310                 315                 320
Arg Tyr Leu Tyr Arg Tyr Tyr Gln Asp Phe Tyr Asp Leu Gln Val Lys
                325                 330                 335
Leu Leu Glu Leu Phe Pro Tyr Glu Ala Gly Arg Ile Glu Asn Ser Lys
            340                 345                 350
Arg Ile Ile Pro Ser Ile Pro Gly Pro Leu Ile Asn Val Asn Asp Ser
        355                 360                 365
Ile Ser Lys Leu Arg Arg Glu Lys Leu Asp Tyr Tyr Leu Ser Asn Leu
    370                 375                 380
Ile Ala Leu Pro Ser His Ile Ser Arg Ser Glu Glu Val Leu Lys Leu
385                 390                 395                 400
Phe Asp Val Leu Asp Asn Gly Phe Asp Arg Glu Thr Asp Ala Ile Asn
                405                 410                 415
```

```
Lys Arg Phe Ser Lys Pro Ile Ser Gln Lys Ser Asn Ser His Gln Asp
                420                 425                 430
Arg Leu Ser Gln Tyr Ser Asn Phe Asn Val Leu Gln Gln Gln Gln Gln
        435                 440                 445
Gln Gln Gln Gln Gln Gln Tyr Ala His His Ser Arg Gly Ser Asp Asn
    450                 455                 460
Ser Pro Thr Asn Glu Ser Ser Gly Ser Asn Leu Ile Asn Ser Ser Ser
465                 470                 475                 480
His Asn Asp Ser Ser Leu Ser Ser Ser Pro Pro Pro Pro Pro Pro Gln
                485                 490                 495
Thr Val Thr Thr Thr Asn Thr Thr Asn Thr Thr Ile Thr Thr Asp Ser
            500                 505                 510
Ser Ser Lys Gln Pro Lys Ala Lys Val Lys Phe Tyr Phe Asp Asp Asp
        515                 520                 525
Ile Phe Val Leu Leu Ile Pro Thr Asn Leu Arg Leu Gln Asp Leu Lys
    530                 535                 540
Ser Lys Leu Phe Lys Arg Leu Glu Leu Asp Ile Thr Tyr Lys Tyr Glu
545                 550                 555                 560
Lys Pro Asp Gln Gln Gln Lys Pro Thr Ser Glu Ser Ile His Leu Phe
                565                 570                 575
Leu Lys Asn Asp Phe Glu Asp Phe Leu Ile Glu Asn Glu Thr Ser Asn
            580                 585                 590
Asn Asn Asn Leu Glu Ile Asp Phe Glu Asn Glu Ile Ile Lys Glu Lys
        595                 600                 605
Leu Gly Glu Phe Glu Val Asn Asp Asp Glu Lys Phe Gln Ser Ile Leu
    610                 615                 620
Phe Asp Lys Cys Lys Leu Met Val Leu Val Tyr
625                 630                 635
```

WE CLAIM:

1. An *in vitro* screening test for compounds to inhibit the biological activity of at least one protein selected from the group consisting of CaCla4p, Cst20p, CaCdc42p and CaBemlp, which comprises:
a) at least one of said proteins; and
b) means to monitor the biological activity of said at least one protein;
thereby compounds are tested for their inhibiting potential.

2. The screening test of claim 1, wherein the inhibition of the interactions between CaCla4p and Ca Cdc42p is determined.

3. The screening test of claim 1, wherein the inhibition of the interactions between Cst20p and CaCdc42p is determined.

4. The screening test of claim 1, wherein the inhibition of the interactions between CaCla4p and CaBemlp is determined.

5. The screening test of claim 1, wherein the inhibition of the interactions between Cst20p and CaBemlp is determined.

6. A method for determining at least one gene involved in filamentous growth associated with virulence, which comprises using one protein selected from the group consisting of CaCla4p, Cst20, CaCdc42p and CaBemlp to determine said gene.

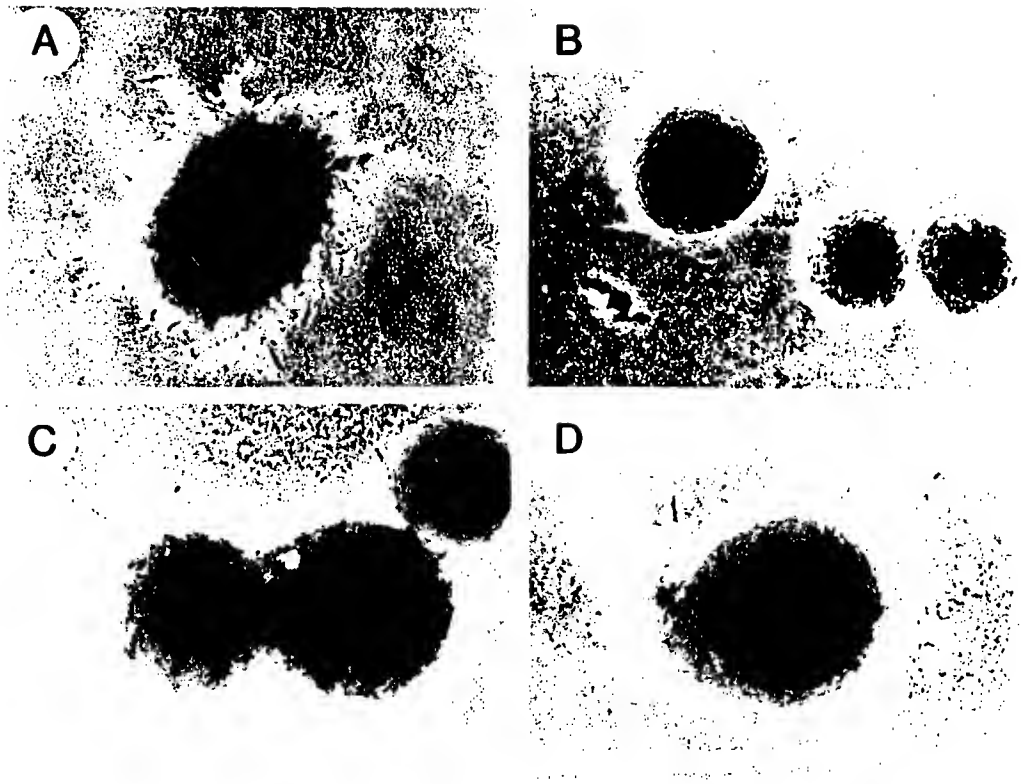


FIG_1

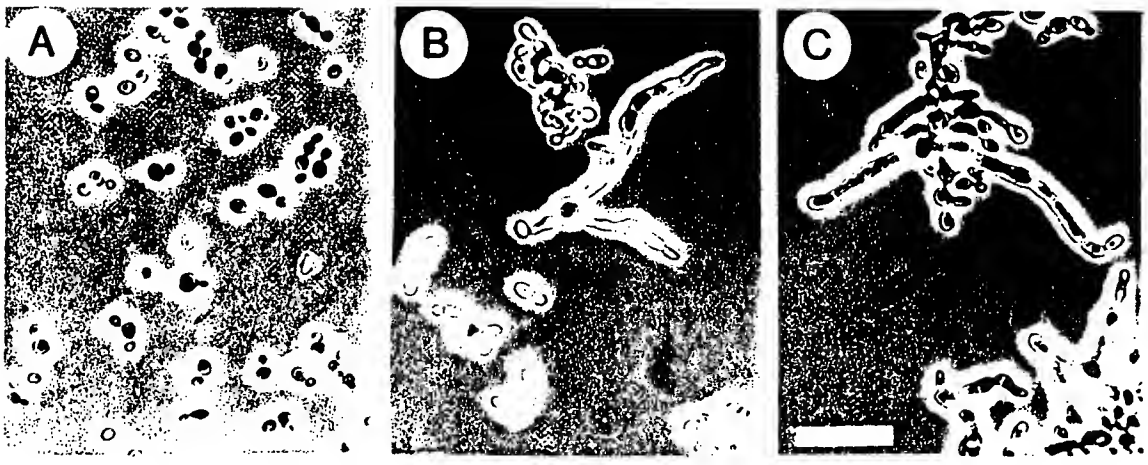


FIG_2

```
-354        GTACCCACTTTACAATCACTTACAAGTCAAATAATTACAACTTGACAATCCTCACTTTAAGTCTAACGTATATACGCGTACACCATCTTATACTCCACATACATATTGGATTCA

-240  ATTTTTATTTTATTGTTTAGTTTATATCCAACCACTGACAATTACCAATAGTTTTCAATTAATATTCACAATTTAACTATTTGTTTGACAGCTGAAAAGAGATAAAAAAAGAATCAAGTG

-120  CTATAACTCACAAGGGCTAGAAATAAGTTTGCAAAAAACAAGTTTTAAAAATAGTAACTGCACTTTTGTTGACTCTTTCACCTCCCCATTGAATTTAACTGAACACAAATAAAGCCTATC

   1  ATGAGCATACTTTCAGAGAACAATCCTACACCAACATCAATAACAGATCCAAATGAGTCTTCTCATCTACACAACCCAGAGTTAAACTCTGGAACGAGGGTTGCTTCTGGACCTGGACCT
   1   M  S  I  L  S  E  N  N  P  T  P  T  S  I  T  D  P  N  E  S  S  H  L  H  N  P  E  L  N  S  G  T  R  V  A  S  G  P  G  P

 121  GGACCTGAAGTTGAATCAACACCACTAGCACCCCCAACTGAGGTCATGAATACTACATCAGCTAATACTTCTTCATTAAGTTTAGGGTCTCCAATGCACGAGAAAATAAAACAATTTGAT
  41   G  P  E  V  E  S  T  P  L  A  P  P  T  E  V  M  N  T  T  S  A  N  T  S  S  L  S  L  G  S  P  M  H  E  K  I  K  Q  F  D

 241  CAAGACGAGGTTGACACTGGGGAAACTAATGATAGGACTATAGAATCTGGATCTAGTGATATTGATGATTCACAACAATCACATAACAACAACAACAACAACAACAACAACAACGAG
  81   Q  D  E  V  D  T  G  E  T  N  D  R  T  I  E  S  G  S  S  D  I  D  D  S  Q  Q  S  H  N  N  N  N  N  N  N  N  N  N  N  E

 361  AGCAATCCAGAATCAAGTGAAGGCGATGATGAAAAAACCCAAGGAATGCCTCCTCGAATGCCAGGGACATTCAATGTGAAAGGTTTGCACCAAGGGGATGATAGTGACAATGAAAAACAG
 121   S  N  P  E  S  S  E  G  D  D  E  K  T  Q  G  M  P  P  R  M  P  G  T  F  N  V  K  G  L  H  Q  G  D  D  S  D  N  E  K  Q

 481  TACACCGAGCTAACTAAATCAATCAATAAACGTACCAGTAAAGATTCGTATTCTCCTGGCACACTTGAAAGTCCCGGTACTCTTAATGCATTGGAAACAAATAATGTCTCACCAGCAGTT
 161   Y  T  E  L  T  K  S  I  N  K  R  T  S  K  D  S  Y  S  P  G  T  L  E  S  P  G  T  L  N  A  L  E  T  N  N  V  S  P  A  V

 601  ATAGAGGAAGAACAACATACACTGTCTTTGGAAGATTTGTCATTGTCCTTACAACACCAAAATGAAAATGCAAGATTATCTGCACCCCGCAGTGCACCGCCACAGGTTCCGACTTCAAAG
 201   I  E  E  E  Q  H  T  S  S  L  E  D  L  S  L  S  L  Q  H  Q  N  E  N  A  R  L  S  A  P  R  S  A  P  P  Q  V  P  T  S  K

 721  ACATCGTCATTTCACGATATGAGTCTGGTTATATCTTCATCAACTTCTGTGCATAAGATACCATCAAATCCAACTTCAACTCGAGGTTCTCATTTATCAAGTTACAAATCTACATTGGAC
 241   T  S  S  F  H  D  M  S  S  V  I  S  S  S  T  S  V  H  K  I  P  S  N  P  T  S  T  R  G  S  H  L  S  S  Y  K  S  T  L  D

 841  CCTGGGAAACCTGCACAAGCAGCAGCACCACCACCACCAGAAATAGACATTGACAATTTATTAACCAAAAGTGAATTGGATCTGGAAACAGACACATTGAGTAGTGCCACAAATTCTCCA
 281   P  G  K  P  A  Q  A  A  A  P  P  P  P  E  I  D  I  D  N  L  L  T  K  S  E  L  D  S  E  T  D  T  L  S  S  A  T  N  S  P

 961  AACCTTTTAAGAAATGATACTTTACAAGGAATTCCAACACGAGATGACGAAAATATTGATGACCTGCCCCGTCAACTATCACAAAATACTAGTGCGACGTCAAGAAATACTTCGGGAACA
 321   N  L  L  R  N  D  T  L  Q  G  I  P  T  R  D  D  E  N  I  D  D  S  P  R  Q  L  S  Q  N  T  S  A  T  S  R  N  T  S  G  T

1081  TCGACTTCTACAGTGGTGAAAAATTCAAGATCTGGTACGTCAAAATCAACCTCAACCTCAACTGCTCATAACCAAACAGCAGCAATTACTCCTATAATCCCGAGTCACAACAAGTTTCAT
 361   S  T  S  T  V  V  K  N  S  R  S  G  T  S  K  S  T  S  T  S  T  A  H  N  Q  T  A  A  I  T  P  I  I  P  S  H  N  K  F  H

1201  CAACAAGTGATAAATACCAATGCAACAAATAGTTCATCTTCACTAGAACCATTGGGGGTTGGCATAAATTCAAATCTGTCTCCTAAAAGTGGGAAAAAGCGGAAAAGTGGAAGTAAAGTC
 401   Q  Q  V  I  N  T  N  A  T  N  S  S  S  S  L  E  P  L  G  V  G  I  N  S  N  S  S  P  K  S  G  K  K  R  K  S  G  S  K  V
```

FIG. 3A

```
321  CGAGGTGTGTTTTCGTCAATGTTTGGGAAAAACAAGTCAACGTCATCATCGTCGTCTTCAAACTCAGGTCTGAATAGCCACTCACAGGAAGTCAATATTAAGATCAGTACTCCATTCAAT
441   R  G  V  F  S  S  M  F  G  K  N  K  S  T  S  S  S  S  S  S  N  S  G  S  N  S  H  S  Q  E  V  N  I  K  I  S  T  P  F  N

441  GCCAAGCACCTTGCCCATGTGGGCATTGATGATAATGGTTCATACACCGGTTTGCCAATAGAGTGGGAAAGATTATTATCTGCTAGTGGTATTACCAAGAAGGAACAACAACAGCACCCA
481   A  K  H  L  A  H  V  G  I  D  D  N  G  S  Y  T  G  L  P  I  E  W  E  R  L  L  S  A  S  G  I  T  K  K  E  Q  Q  Q  H  P

561  CAAGCAGTGATGGATATAGTGGCGTTTTATCAAGATACAAGTGAAAACCCTGATGACGCTGCATTTAAAAAGTTTCATTTTGATAATAATAAAAGTAGTTCGAGTGGTTGGTCTAATGAA
521   Q  A  V  M  D  I  V  A  F  Y  Q  D  T  S  E  N  P  D  D  A  A  F  K  K  F  H  F  D  N  N  K  S  S  S  S  G  W  S  N  E

681  AATACTCCACCAGCAACACCGGGTGGGAGTAACAGTGGCAGTGGCAGTGGTGGCGGTGGCGCTCCTTCAAGTCCCCATCGTACACCTCCTTCATCGATCATTGAAAAAAACAACGTTGAA
561   N  T  P  P  A  T  P  G  G  S  N  S  G  S  G  S  G  G  G  G  A  P  S  S  P  H  R  T  P  P  S  S  I  I  E  K  N  N  V  E

801  CAAAAAGTGATTACCCCATCTCAGTCAATGCCAACAAAGACAGAGAGTAAACAGCTGGAAAACCAGCACCCACATGAAGATAATGCTACTCAGTATACACCAAGAACACCAACATCCCAT
601   Q  K  V  I  T  P  S  Q  S  M  P  T  K  T  E  S  K  Q  S  E  N  Q  H  P  H  E  D  N  A  T  Q  Y  T  P  R  T  P  T  S  H

921  GTACAAGAGGGTCAATTTATTCCAAGTAGACCAGCTCCGAAACCACCATCAACACCGCTTTCTTCCATGAGTGTGTCACATAAAACACCTTCTTCGCAATCATTACCAAGGAGTGATTCA
641   V  Q  E  G  Q  F  I  P  S  R  P  A  P  K  P  P  S  T  P  L  S  S  M  S  V  S  H  K  T  P  S  S  Q  S  L  P  R  S  D  S

041  CAATCCGATATTCGTTCTTCAACCCCTAAATCACATCAAGATGTTTCGCCAAGCAAGATCAAAATTCGTTCAATTTCGTCAAAATCATTAAAGTCAATGCGGTCTAGAAAAAGTGGGGAT
681   Q  S  D  I  R  S  S  T  P  K  S  H  Q  D  V  S  P  S  K  I  K  I  R  S  I  S  S  K  S  L  K  S  M  R  S  R  K  S  G  D

161  AAGTTTACTCATATTGCACCTGCTCCTCCACCACCATCATTACCTTCAATTCCTAAATCAAAGTCGCATTCGGCATCTTTGTCAAGTCAATTGAGACCAGCAACAAATGGATCAACAACT
721   K  F  T  H  I  A  P  A  P  P  P  P  S  L  P  S  I  P  K  S  K  S  H  S  A  S  L  S  S  Q  L  R  P  A  T  N  G  S  T  T

281  GCCCCTATTCCAGCAAGTGCCGCGTTTGGTGGTGAGAATAATGCTTTACCAAAACAAAGAATAAATGAGTTCAAGGCTCATAGAGCACCTCCACCACCTCCACTGGCACCACCTGCACCA
761   A  P  I  P  A  S  A  A  F  G  G  E  N  N  A  L  P  K  Q  R  I  N  E  F  K  A  H  R  A  P  P  P  P  P  S  A  P  P  A  P

401  CCTGTGCCTCCTGCTCCACCAGCCAATTTATTATCGGAACAGACTTCTGAGATACCTCAACAACGTACTGCTCCTCTGCAAGCATTAGCTGATGTTACTGCCCCAACTAATATTTATGAA
801   P  V  P  P  A  P  P  A  N  L  L  S  E  Q  T  S  E  I  P  Q  Q  R  T  A  P  S  Q  A  L  A  D  V  T  A  P  T  N  I  Y  E

521  ATTCAACAAACTAAATATCAGGAAGCACAACAGAAATTACGTGAGAAGAAGGCTAGAGAACTTGAAGAAATACAAAGACTACGAGAGAAGAATGAAAGACAAAATAGACAACAGGAGACT
841   I  Q  Q  T  K  Y  Q  E  A  Q  Q  K  L  R  E  K  K  A  R  E  L  E  E  I  Q  R  L  R  E  K  N  E  R  Q  N  R  Q  Q  E  T
```


Fig. 3B

```
41   GGGCAAAATAATGCTGACACGGCTAGCGGTGGTAGTAATATTGCTCCACCAGTACCTGTACCAAATAAAAAACCGCCTTCTGGATCTGGTGGTGGCCGTGATGCCAAACAAGCAGCTTTG
81     G  Q  N  N  A  D  T  A  S  G  G  S  N  I  A  P  P  V  P  V  P  N  K  K  P  P  S  G  S  G  G  G  R  D  A  K  Q  A  A  L

61   ATAGCCCAAAAGAAACGAGAAGAAAAGAAACGTAAAAACTTACAAATTATTGCCAAATTAAAGACAATTTGTAATCCTGGAGATCCAAATGAATTATATGTTGATTTAGTTAAAATTGGT
21     I  A  Q  K  K  R  E  E  K  K  R  K  N  L  Q  I  I  A  K  L  K  T  I  C  N  P  G  D  P  N  E  L  Y  V  D  L  V  K  I  G

81   CAAGGTGCCTCCGGTGGAGTTTTCCTTGCTCATGATGTTCGTGATAAATCCAATATTGTTGCCATAAAACAAATGAATTTAGAACAACAACCTAAAAAAGAATTAATTATTAATGAAATT
61     Q  G  A  S  G  G  V  F  L  A  H  D  V  R  D  K  S  N  I  V  A  I  K  Q  M  N  L  E  Q  Q  P  K  K  E  L  I  I  N  E  I

01   TTGGTTATGAAAGGTAGTCTGCATCCCAATATTGTCAATTTTATTGATTCATATCTTTTAAAAGGTGATTTATGGGTGATTATGGAATATATGGAAGGTGGATCCCTTACCGATATAGTG
01     L  V  M  K  G  S  S  H  P  N  I  V  N  F  I  D  S  Y  L  L  K  G  D  L  W  V  I  M  E  Y  M  E  G  G  S  L  T  D  I  V

21   ACTCATAGTGTTATGACCGAAGGTCAAATTGGAGTTGTATGTCGTGAAACTTTGAAAGGTCTTAAATTTTTACATTCCAAAGGGGTTATCCATCGTGATATTAAATCCGATAATATTTTA
41     T  H  S  V  M  T  E  G  Q  I  G  V  V  C  R  E  T  L  K  G  L  K  F  L  H  S  K  G  V  I  H  R  D  I  K  S  D  N  I  L

41   TTAAATATGGATGGTAACATCAAGATCACTGATTTTGGGTTTTGTGCTCAAATCAATGAAATCAATCTGAAACGTATCACTATGGTGGGTACACCATATTGGATGGCACCAGAAATTGTT
81     L  N  M  D  G  N  I  K  I  T  D  F  G  F  C  A  Q  I  N  E  I  N  S  K  R  I  T  M  V  G  T  P  Y  W  M  A  P  E  I  V

61   TCACGTAAAGAGTATGGTCCAAAAGTTGATGTTTGGTCATTAGGTATCATGATTATAGAAATGTTAGAAGGTGAACCACCATATTTGAATGAAACTCCATTGAGGGCATTATATCTTATT
21     S  R  K  E  Y  G  P  K  V  D  V  W  S  L  G  I  M  I  I  E  M  L  E  G  E  P  P  Y  L  N  E  T  P  L  R  A  L  Y  L  I

81   GCAACTAATGGTACACCAAAATTAAAAGATCCTGAATCTTTAAGTTATGATATTAGAAAATTTTTGGCATGGTGTTTACAAGTTGACTTTAATAAAAGAGCTGATGCTGATGAATTATTA
61     A  T  N  G  T  P  K  L  K  D  P  E  S  L  S  Y  D  I  R  K  F  L  A  W  C  L  Q  V  D  F  N  K  R  A  D  A  D  E  L  L

01   CATGATAATTTTATTACTGAATGTGATGATGTATCGTCGTTAAGTCCATTAGTGAAAATTGCTCGATTGAAAAAAATGAGTGAATCTGATTAATGAATGGTGGAGTTATCCTAGAAATAA
01     H  D  N  F  I  T  E  C  D  D  V  S  S  L  S  P  L  V  K  I  A  R  L  K  K  M  S  E  S  D  -

721  AAACATTTAAAAAAAAAGAAGAAGAACAACAAGAACCCTAAATTCTACTGCTGTCAATATATTGGCTAATTTCCATTCTCGTTTCTATTTCTATTTCGTTTTTATTCTTTGAATTATTAT

341  TGTTAGTGGTAGAGATTTTTACTAGTATATTTTTTTTATTCATTTTTATATTTGTATTTATATATATATTTTTCATTTAGTATTTACTTACACTGCAGTATCTTTCTTTTCTGTGTAGAT

961  GATATGTAGTAATAAGTTAACTTGTTCAAGACAGTGAATGGAAATATATTATAGCTTGACTATATAAGGTGGAGAGCTGTAATTGGCTTTCCGTATAGAAAAGTCTTGAACAAACGTTAC

4081  CAGATTTCTGCTATTCTTATTTGGTACGATTCGGGCGTATGATAGGTTTATTGAGCTC
```

4/16

Fig. 3C

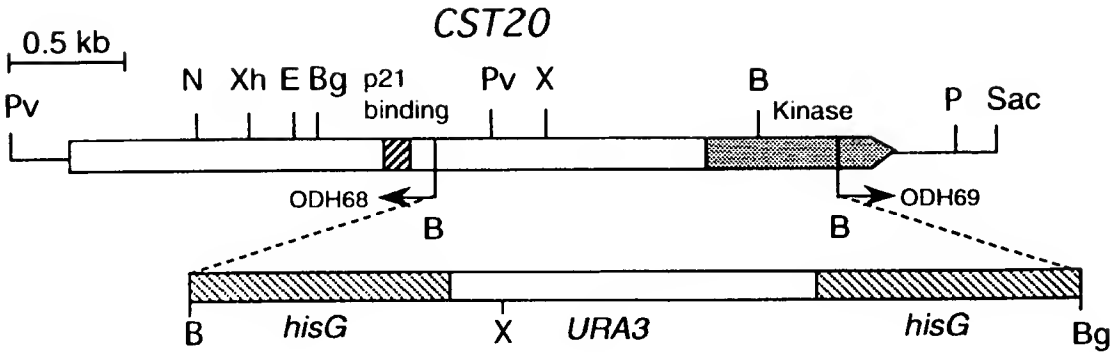

CST20
0.5 kb
Pv
N Xh E Bg p21 binding Pv X
B Kinase
P Sac
ODH68
ODH69
B
B
B hisG X URA3 hisG Bg

Fig 4A

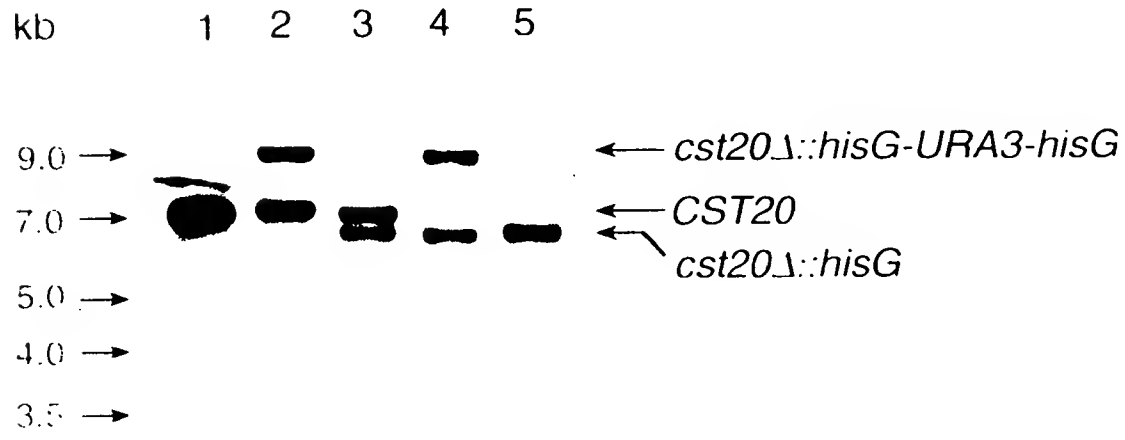
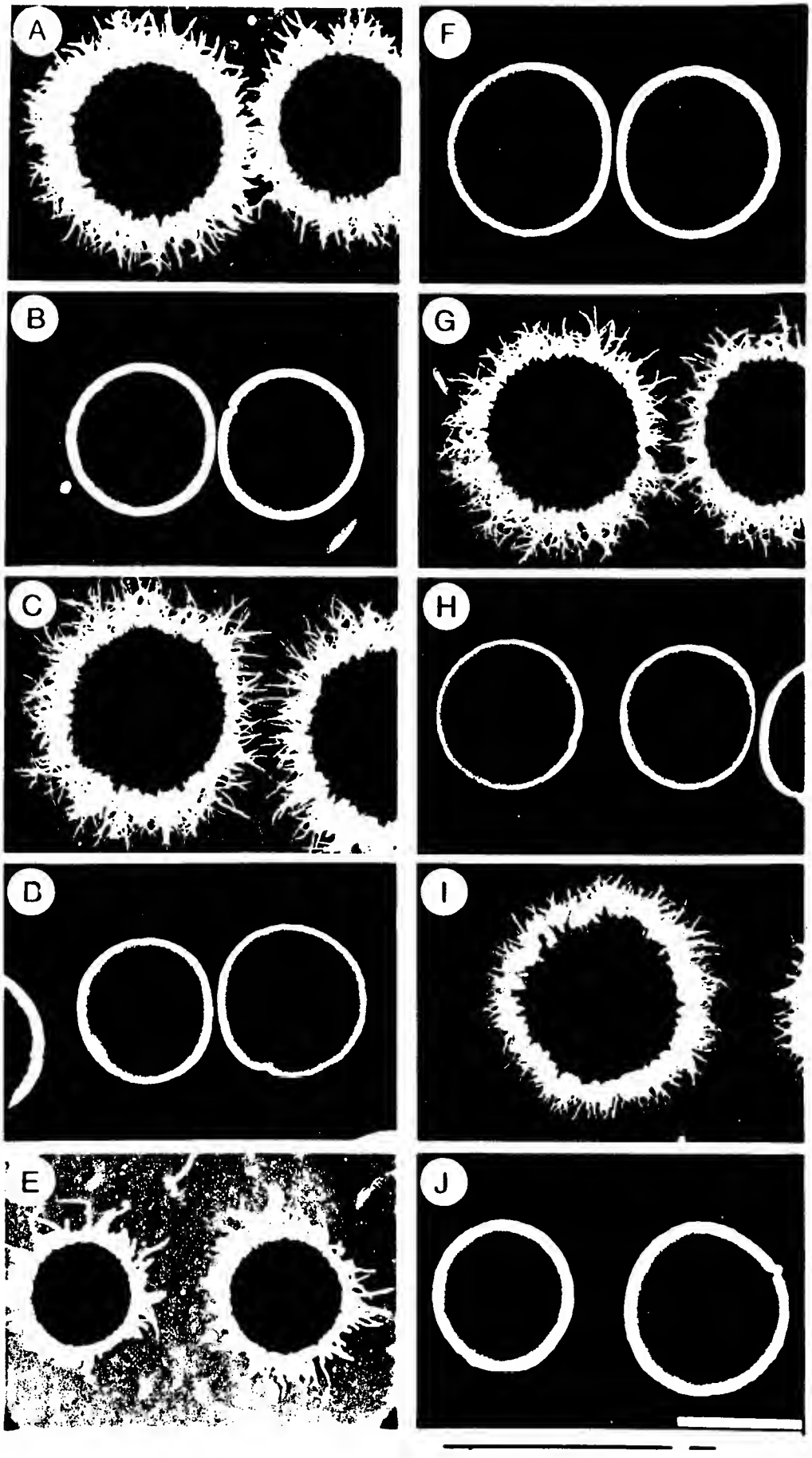

kb
1 2 3 4 5
9.0
7.0
5.0
4.0
3.5
cst20Δ::hisG-URA3-hisG
CST20
cst20Δ::hisG

Fig 4B


A
B
C
D
E
F
G
H
I
J

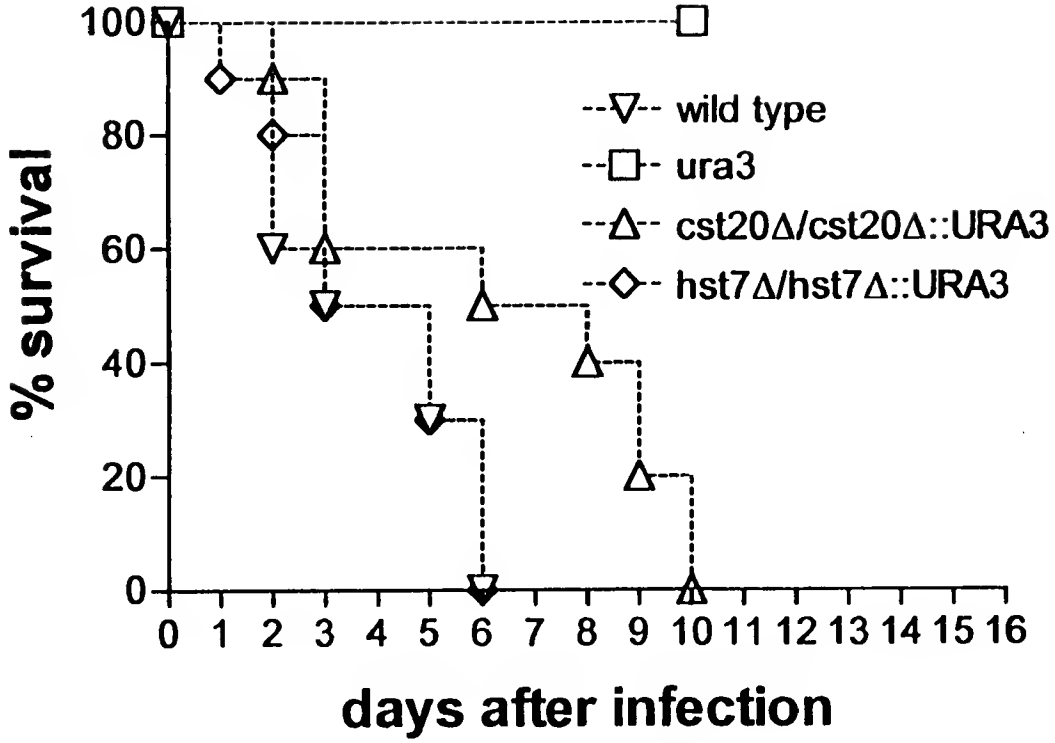

100
80
60
40
20
0
% survival
wild type
ura3
cst20Δ/cst20Δ::URA3
hst7Δ/hst7Δ::URA3
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
days after infection

Fig 6A

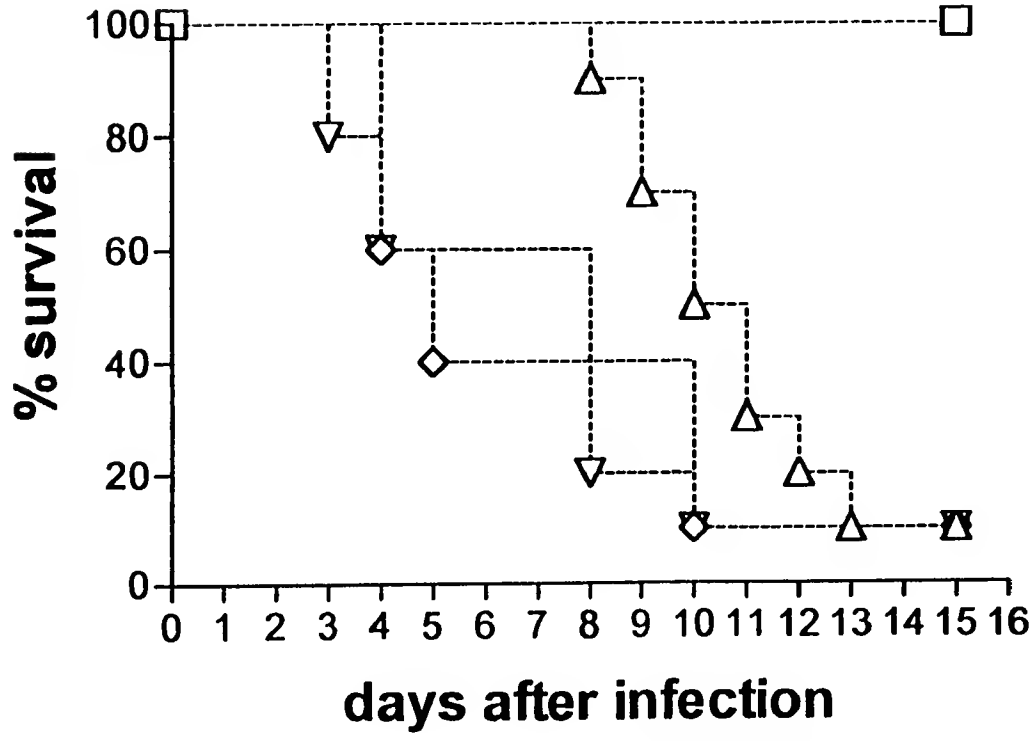

100
80
60
40
20
0
% survival
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
days after infection

Fig 6B

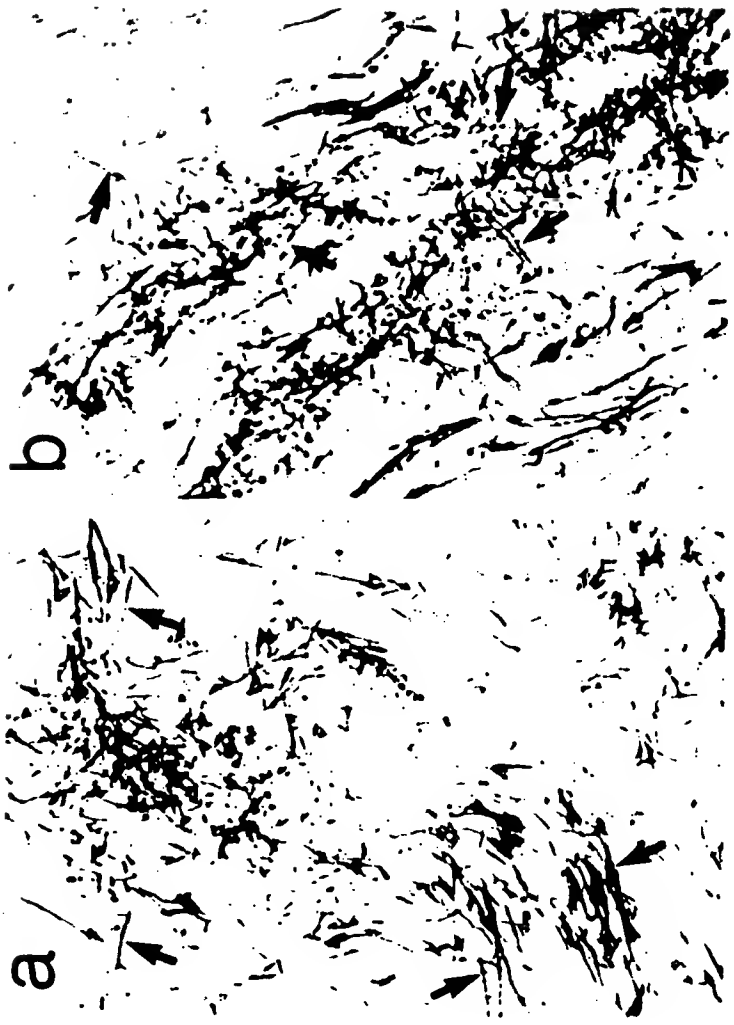


FIG - 8C

```
1   GAATTCTTTTTAGAAGAGAAAGAAAAAATTCCCAAAAAAAAAAGATTTCATTTAATTCCACGGGAACATTG

0   ATTACAACCACGTCAACAGTTTCCCTTTTATATTGAAATCAACATTCAATTTTGTCTTTTTTTTTTTTTTCATTGATTTTTCCCCAATCTTTTTATCTTCATATTAATATTGGATATCAAT

0   TACTAATACTGTCAGGGATAGTTTAGTAAATATTTACATTCTCCATTCAATCCTAAATTTTTTTTATATAGCTAGTTTTTGGTTGAAAAAAAAAAAAATAGGGGGAAGGAAGTTTTTTTT

0   TCTATTTATTTAATTGTTTTGATTCCAACCATATTGTATATTTGTCTTGTCAGTTATATTACTTTCTTGTTACTTAATTATTAATTATTTGCTATATTATTGAATTGAATCCTCAAAAGA

1   ATGACAAGTATTTATACATCAGATTTGAAAAACCATAGACGTGCGCCACCTCCACCAAATGGGGCAGCTGGCTCTGGCTCAGGTTCTGGCTCAGGTTCTGGTTCTGGTTCTGGCAGTTTG
1    M  T  S  I  Y  T  S  D  L  K  N  H  R  R  A  P  P  P  P  N  G  A  A  G  S  G  S  G  S  G  S  G  S  G  S  G  S  G  S  L

:1  GCTAATATTGTTACCAGTTCTAATAGTCTTGGCGTAACAGCAAATCAAACCAAACCTATTCAATTAAATATAAATTCTAGCAAACGTCAATCAGGTTGGGTTCATGTTAAAGATGATGGT
:1   A  N  I  V  T  S  S  N  S  L  G  V  T  A  N  Q  T  K  P  I  Q  L  N  I  N  S  S  K  R  Q  S  G  W  V  H  V  K  D  D  G

.1  ATTTTCACATCATTTAGATGGAACAAACGGTTTATGGTTATTAATGATAAAACTTTAAACTTTTATAAACAAGAACCATATTCTAGTGATGGTAATTCCAATTCTAATACCCCTGATTTA
31   I  F  T  S  F  R  W  N  K  R  F  M  V  I  N  D  K  T  L  N  F  Y  K  Q  E  P  Y  S  S  D  G  N  S  N  S  N  T  P  D  L

;1  TCATTCCCACTATATTTAATTAATAATATTAATTTGAAACCAAACTCCGGGTATAGCAAAACTTCACAATCATTTGAAATTGTTCCCAAAAACAATAATAAATCAATTTTGATTTCTGTT
21   S  F  P  L  Y  L  I  N  N  I  N  L  K  P  N  S  G  Y  S  K  T  S  Q  S  F  E  I  V  P  K  N  N  N  K  S  I  L  I  S  V

31  AAAACCAATAATGATTATTTGGATTGGCTAGATGCATTCACCACAAAATGTCCTTTAGTACAAATTGGTGAAAATAATAGTGGTGTATCAAGTAGTCACCCTCATTTACAAATTCAACAT
;1   K  T  N  N  D  Y  L  D  W  L  D  A  F  T  T  K  C  P  L  V  Q  I  G  E  N  N  S  G  V  S  S  S  H  P  H  L  Q  I  Q  H

)1  TTAACCAATGGTTCCTTGAACGGCAACTCATCTTCATCACCAACATCTGGATTATTATCTTCTTCAGTGCTAACTGGAGGTAATTCTGGCGTTTCTGGTCCTATTAATTTCACTCATAAA
)1   L  T  N  G  S  L  N  G  N  S  S  S  S  P  T  S  G  L  L  S  S  S  V  L  T  G  G  N  S  G  V  S  G  P  I  N  F  T  H  K

21  GTACACGTGGGATTTGATCCTGCCAGTGGTAATTTTACTGGATTACCAGACACTTGGAAAAGTTTATTACAACATTCGAAAATCACTAATGAGGATTGGAAAAAAGATCCTGTTGCTGTT
11   V  H  V  G  F  D  P  A  S  G  N  F  T  G  L  P  D  T  W  K  S  L  L  Q  H  S  K  I  T  N  E  D  W  K  K  D  P  V  A  V

11  ATTGAAGTTTTAGAATTTTATTCCGATATAAATGGAGGTAATTCAGCTGCTGGAACTCCAATTGGATCACCCATGATCAATTCCAAAACCAACAATAATAATAATGACCCTAACAATTAC
31   I  E  V  L  E  F  Y  S  D  I  N  G  G  N  S  A  A  G  T  P  I  G  S  P  M  I  N  S  K  T  N  N  N  N  N  D  P  N  N  Y

51  TCATCAACCAAAAACAATGTCCAAGAGGCAAATTTACAAGAATGGGTAAAACCTCCAGCAAAATCTACTGTCTCACAATTCAAACCTAGTCGAGCTGCACCAAAACCACCAACTCCATAT
21   S  S  T  K  N  N  V  Q  E  A  N  L  Q  E  W  V  K  P  P  A  K  S  T  V  S  Q  F  K  P  S  R  A  A  P  K  P  P  T  P  Y

31  CATTTGACACAACTAAATGGCTCTTCCCACCAACATACATCATCATCAGGCTCATTACCTAGTTCTGGTAATAATAATAATAATAACAGCACTAACAATAATAATACTAAAAACGTTTCA
51   H  L  T  Q  L  N  G  S  S  H  Q  H  T  S  S  S  G  S  L  P  S  S  G  N  N  N  N  N  N  S  T  N  N  N  N  T  K  N  V  S

)1  CCATTGAATAATTTGATGAATAAATCTGAACTTATTCCTGCTAGAAGAGCTCCACCACCTCCAACAAGTGGCACATCTTCAGATACATATTCTAATAAGAATCATCAAGATAGATCTGGA
)1   P  L  N  N  L  M  N  K  S  E  L  I  P  A  R  R  A  P  P  P  P  T  S  G  T  S  S  D  T  Y  S  N  K  N  H  Q  D  R  S  G
```

FIG. 7A

```
321  TATGAACAACAACGTCAACAACGTACTGACTCATCACAACAACAACAACAACAAAAGCAACATCAATATCAACAGAAATCCCAACAACAACAACAACAACCACAACAACCATTATCTCTG
441   Y  E  Q  Q  R  Q  Q  R  T  D  S  S  Q  Q  Q  Q  Q  Q  K  Q  H  Q  Y  Q  Q  K  S  Q  Q  Q  Q  Q  Q  P  Q  Q  P  L  S  S

441  CATCAAGGTGGGACTTCGCATATTCCGAAACAAGTACCTCCTACATTACCATCATCTGGACCACCCACTCAGGCTGCTTCAGGAAAATCAATGCCATCTAAAATTCATCCTGATCTTAAG
481   H  Q  G  G  T  S  H  I  P  K  Q  V  P  P  T  L  P  S  S  G  P  P  T  Q  A  A  S  G  K  S  M  P  S  K  I  H  P  D  L  K

561  ATTCAACAAGGCACAAATAATTATATTAAGAGTAGCGGTACTGATGCTAATCAAGTCGATGGTGATGCTAAACAATTTATTAAACCATTTAATTTACAACTGAAAAAGAGTCAGCAACAA
521   I  Q  Q  G  T  N  N  Y  I  K  S  S  G  T  D  A  N  Q  V  D  G  D  A  K  Q  F  I  K  P  F  N  L  Q  S  K  K  S  Q  Q  Q

681  TTGGCATCAAAACAACCGTCACCACCTTCATCTCAACAACAGCAACAAAAACCTATGACATCACATGGATTAATGGGTACATCACATTCAGTTACTAAACCATTGAATCCAGTCAATGAT
561   L  A  S  K  Q  P  S  P  P  S  S  Q  Q  Q  Q  Q  K  P  M  T  S  H  G  L  M  G  T  S  H  S  V  T  K  P  L  N  P  V  N  D

801  CCAATCAAACCATTAAATTTGAAATCATCTAAATCCAAAGAAGCATTAAATGAAACTCTGGGGGGTGCTGAAAACACCATCACCTACAGATAAATCAAATAAACCAACTGCACCTGCTAGT
601   P  I  K  P  L  N  L  K  S  S  K  S  K  E  A  L  N  E  T  S  G  V  S  K  T  P  S  P  T  D  K  S  N  K  P  T  A  P  A  S

921  GGTCCTGCAGTGACCAAAACAGCTAAACAACTCAAGAAGGAACGAGAAAGATTGAATGATTTACAAATCATTGCTAAATTGAAAACAGTGGTTAATAATCAAGATCCTAAACCATTATTT
641   G  P  A  V  T  K  T  A  K  Q  L  K  K  E  R  E  R  L  N  D  L  Q  I  I  A  K  L  K  T  V  V  N  N  Q  D  P  K  P  L  F

1041  AGAATTGTTGAAAAAGCTGGTCAAGGTGCTAGTGGGAATGTTTATTTGGCGGAAATGATCAAAGATAATAATCGAAAGATTGCGATTAAACAAATGGATCTTGATGCTCAACCCCGTAAA
681   R  I  V  E  K  A  G  Q  G  A  S  G  N  V  Y  L  A  E  M  I  K  D  N  N  R  K  I  A  I  K  Q  M  D  L  D  A  Q  P  R  K

1161  GAGTTAATAATAAATGAAATCTTGGTTATGAAAGATAGTCAACATAAAAATATTGTTAATTTTTTGGATTCTTATTTAATTGGTGATAATGAATTATGGGTAATTATGGAATATATGCAA
721   E  L  I  I  N  E  I  L  V  M  K  D  S  Q  H  K  N  I  V  N  F  L  D  S  Y  L  I  G  D  N  E  L  W  V  I  M  E  Y  M  Q

1281  GGTGGTTCATTAACGGAAATCATTGAAAATAATGATTTTAAATTGAATGAAAAACAAATTGCCACTATATGTTTTGAAACCTTAAAGGGGTTACAACATTTACATAAAAAACATATTATT
761   G  G  S  L  T  E  I  I  E  N  N  D  F  K  L  N  E  K  Q  I  A  T  I  C  F  E  T  L  K  G  L  Q  H  L  H  K  K  H  I  I

1401  CATCGTGATATTAAATCCGATAATGTTTTATTAGATGCATATGGTAATGTTAAAATCACTGATTTTGGATTTTGTGCTAAATTAACTGATCAAAGAAATAAACGTGCCACAATGGTGGGG
801   H  R  D  I  K  S  D  N  V  L  L  D  A  Y  G  N  V  K  I  T  D  F  G  F  C  A  K  L  T  D  Q  R  N  K  R  A  T  M  V  G

1521  ACACCATATTGGATGGCACCTGAAGTGGTTAAACAAAAGGAATATGATGAAAAAGTTGATGTTTGGTCATTGGGGATTATGACTATTGAAATGATTGAAGGAGAACCACCTTATTTGAAT
841   T  P  Y  W  M  A  P  E  V  V  K  Q  K  E  Y  D  E  K  V  D  V  W  S  L  G  I  M  T  I  E  M  I  E  G  E  P  P  Y  L  N

1641  GAAGAACCATTAAAAGCTTTATATCTTATAGCTACTAATGGTACACCAAAATTGAAAAAACCCGAATTATTATCGAATTCAATTAAAAAATTCTTATCAATTTGTCTTTGTGTTGATGTT
881   E  E  P  L  K  A  L  Y  L  I  A  T  N  G  T  P  K  L  K  K  P  E  L  L  S  N  S  I  K  K  F  L  S  I  C  L  C  V  D  V

1761  AGATATCGTGCTAGTACTGATGAATTATTAGAACATTCATTTATTCAACATAAATCAGGGAAAATTGAAGAATTGGCACCATTATTAGAATGGAAAAAACAACAACAAAAGCATCAACAG
921   R  Y  R  A  S  T  D  E  L  L  E  H  S  F  I  Q  H  K  S  G  K  I  E  E  L  A  P  L  L  E  W  K  K  Q  Q  Q  K  H  Q  Q

1881  CATAAACAAGAAACACTGGATACAGGATTTGCATAGAGATTGAATATAGCCGTAGAAAACTGGTACTTTGGTTTTGGTATAATATATTTATGTGATGTGTTGTGTGTATGGTTAGTTTAG
961   H  K  Q  E  T  S  D  T  G  F  A  *

3001  ATTTGGATTTTTAGTTTTTTAGAGTTTAGTTTTTCAATTTTTAGTTTTAGAGACAATATTCTAGA  3065
```

10/16

Fig 7B

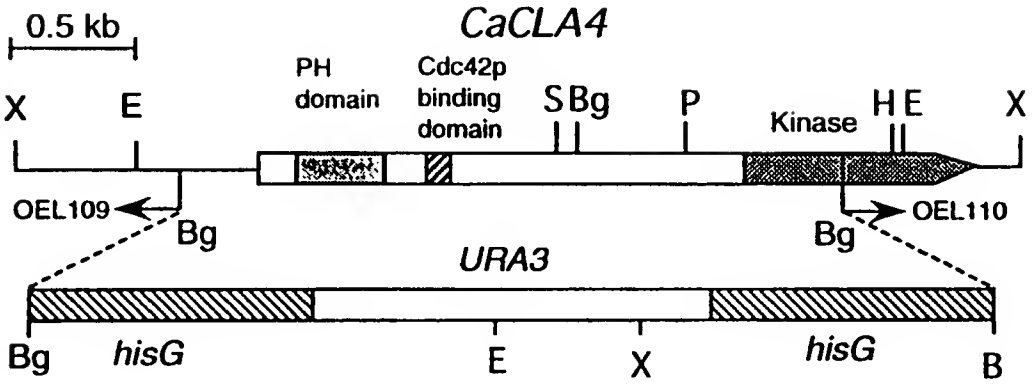

0.5 kb
CaCLA4
PH domain
Cdc42p binding domain
X
E
S
Bg
P
Kinase
H
E
X
OEL109
Bg
Bg
OEL110
URA3
Bg
hisG
E
X
hisG
B

Fig 8A

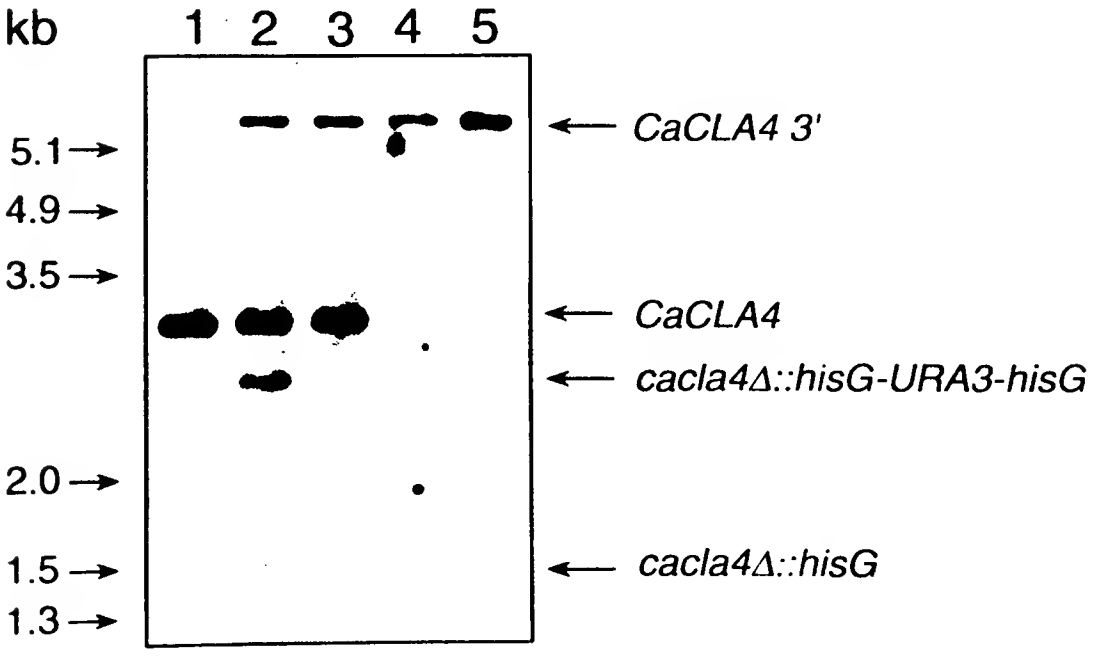

kb
1
2
3
4
5
5.1
4.9
3.5
2.0
1.5
1.3
CaCLA4 3'
CaCLA4
cacla4Δ::hisG-URA3-hisG
cacla4Δ::hisG

Fig 8B

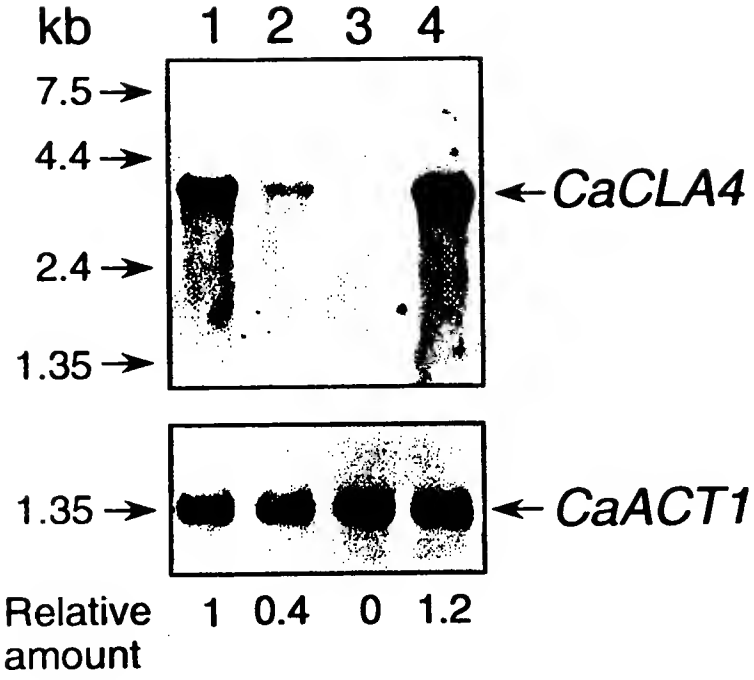

kb
1 2 3 4
7.5
4.4
2.4
1.35
CaCLA4
1.35
CaACT1
Relative amount
1 0.4 0 1.2

Fig. 8C

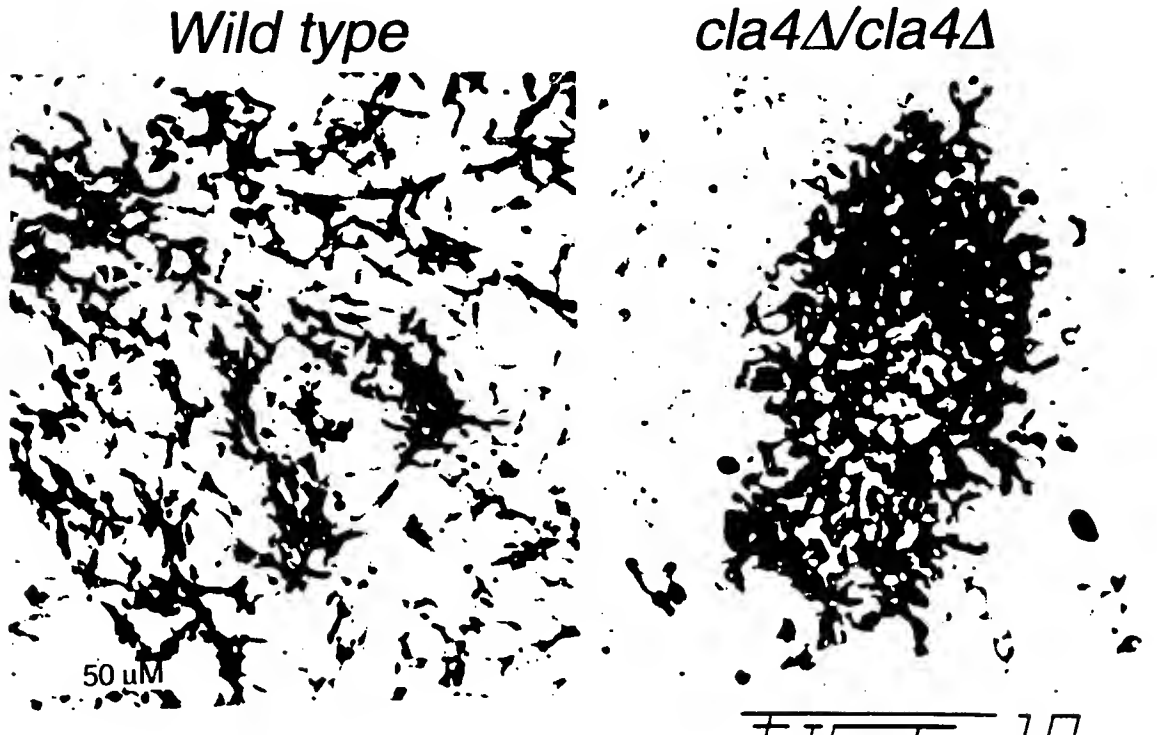

Wild type
cla4Δ/cla4Δ
50 uM

Fig. 10

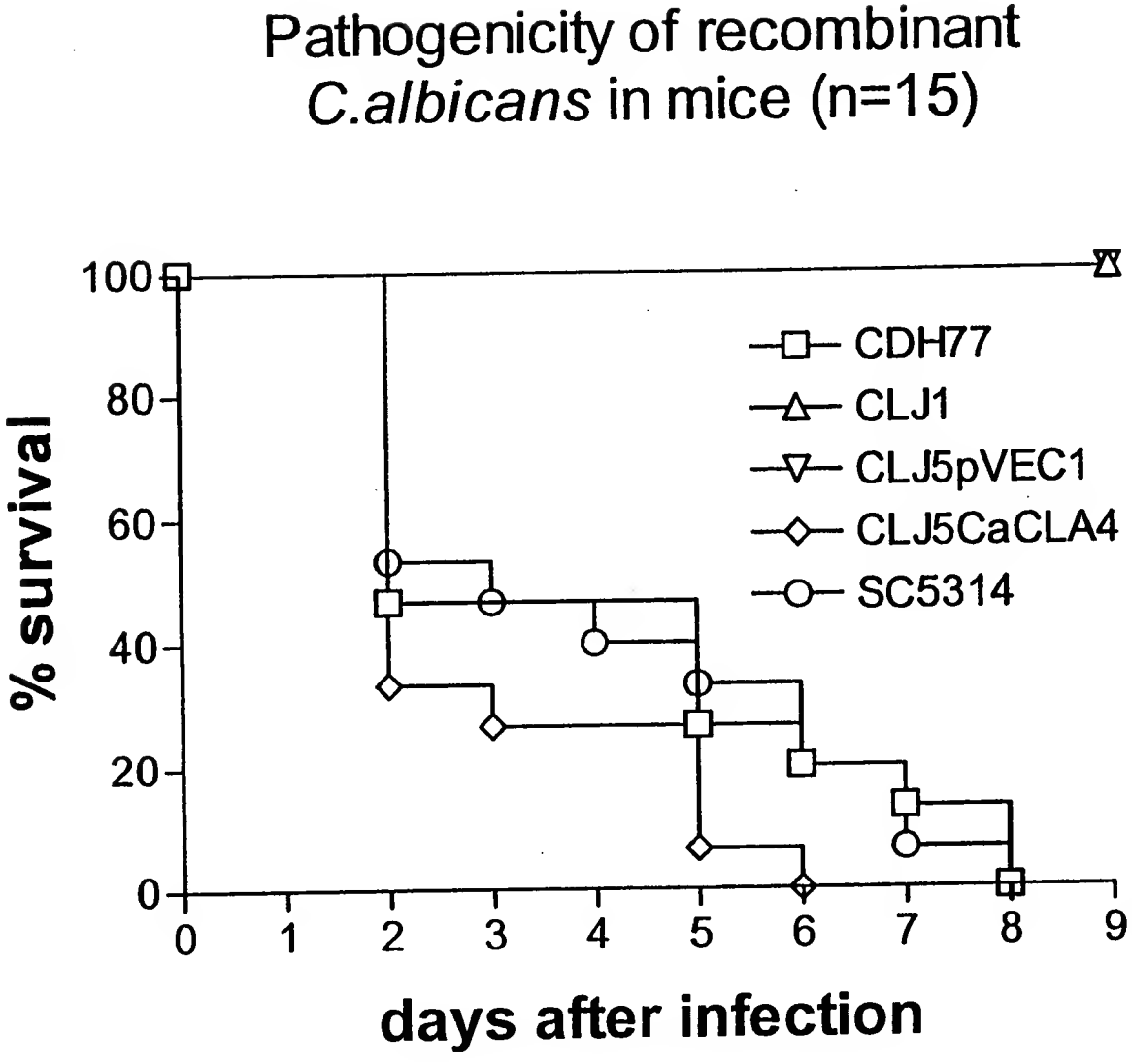

Pathogenicity of recombinant
C.albicans in mice (n=15)
% survival
100
80
60
40
20
0
0
1
2
3
4
5
6
7
8
9
days after infection
CDH77
CLJ1
CLJ5pVEC1
CLJ5CaCLA4
SC5314

Fig 9

```
70  CAACCAAACCAACTTTCATCCTTCTACCAATATCTTCAACAAAAGTTTTATTCAATACTATTTTAAAAATAACAGTGTTACTCGTTCATT

30  TGATTTGTTAATAAGACTGATTTACCCACTTTTTAGTTCCTATAATCATACAGATTTCTCGTCCTAAATCTATTTTTATTGTTATTTTTA

90  CTTTAGTTTTCACTTTTGCTTTCAGTTTTTTCTTTTTTTTAGCACAAGAGAAAAGTATTCAGCTCATAAATAATTAATATATCCATATATC

 1  ATGCAAACTATAAAATGTGTTGTTGTCGGTGATGGTGCCGTTGGTAAAACTTGCTTATTAATCTCGTATACCACTAGTAAATTTCCAGCT
 1   M  Q  T  I  K  C  V  V  V  G  D  G  A  V  G  K  T  C  L  L  I  S  Y  T  T  S  K  F  P  A

91  GATTATGTTCCTACTGTTTTTGATAATTATGCTGTAACCGTGATGATAGGAGACGAACCATTTACCTTGGGATTATTTGATACTGCTGGT
31   D  Y  V  P  T  V  F  D  N  Y  A  V  T  V  M  I  G  D  E  P  F  T  L  G  L  F  D  T  A  G

81  CAAGAAGATTACGACAGATTAAGGCCTTTGTCATATCCATCGACTGATGTATTCCTTGTTTGTTTTTCCGTCATTTCTCCCGCTTCGTTT
61   Q  E  D  Y  D  R  L  R  P  L  S  Y  P  S  T  D  V  F  L  V  C  F  S  V  I  S  P  A  S  F

71  GAAAATGTTAAAGAAAAATGGTTCCCAGAAGTTCATCACCATTGTCCCGGTGTGCCAATAATTATTGTCGGTACCCAAACTGATTTACGA
91   E  N  V  K  E  K  W  F  P  E  V  H  H  H  C  P  G  V  P  I  I  I  V  G  T  Q  T  D  L  R

61  AACGATGATGTTATTTTACAGAGATTGCACAGACAAAAATTGTCCCCAATCACCCAGGAACAGGGTGAAAAATTGGCTAAGGAATTGAGA
21   N  D  D  V  I  L  Q  R  L  H  R  Q  K  L  S  P  I  T  Q  E  Q  G  E  K  L  A  K  E  L  R

51  GCTGTCAAGTATGTTGAGTGTTCTGCATTGACTCAAAGAGGATTGAAAACAGTGTTTGACGAGGCTATAGTAGCTGCATTAGAACCTCCT
51   A  V  K  Y  V  E  C  S  A  L  T  Q  R  G  L  K  T  V  F  D  E  A  I  V  A  A  L  E  P  P

41  GTAATTAAAAAATCGAAAAAGTGTACTATTTTATAGGTCGGCGATACTAGAAGATAGAGGATATTGGAAATAGGGCATACATGAGATATT
81   V  I  K  K  S  K  K  C  T  I  L  -

31  GAATATCTATCATTAAATATATAATTAGTTTTTTTCTAAAACCTATCTTTAGGTTTGATCTCGTTTGATGTGTTGGGCTGTTTCGCAAAA

21  CAGTGTTCCAATCAATAAAAAGATGTGTGTAAGACTCTAGA  761
```


Fig. 11

```
291                                                                  AAGCTTGTTTCTTATCTCCT

270  TAGTATATTGTTTTACAACACCACATACACATACACATATAGCCTTCATTAGCCTTCATTTTGACATATTTCAATAACAATCAAGAACTA

180  CAAGTCATAACTGACACACATATAATATCTTAATTGTTATTATAAATTTATTCTTGATTAGATTTTAGACGGGCAGAAACAAAAACGGAA

-90  AATCCAACTCATCCCCGATAACTACACACATCTATATTAAATCATCTATTAGTCTATCAGTTATATCTCCCTCCCCTTTTCTTCTAACAA

  1  ATGATTAAGACGTTTCGGAAAAGTAAAAGACTGTCGAGTAATTCAAGTTCACCCAAGAAAACAATATCTCGAGTATCATCAACTTCAAGT
  1   M  I  K  T  F  R  K  S  K  R  S  S  S  N  S  S  S  P  K  K  T  I  S  R  V  S  S  T  S  S

 91  AATCAAACATCTCATGATGGAATATTACAATCACCTAAAAAAGTCATTAGAGCTCTATATGATTATGAACCTCAAGGTCCTGGAGAATTG
 31   N  Q  T  S  H  D  G  I  L  Q  S  P  K  K  V  I  R  A  L  Y  D  Y  E  P  Q  G  P  G  E  L

181  AAATTTTTCAAAGGAGATTTTTTCCATGTATTAAATGATGTTGATGATGAATTACATAAAGAAGCGGAAGCTAATGGATGGATAGAAGCA
 61   K  F  F  K  G  D  F  F  H  V  L  N  D  V  D  D  E  L  H  K  E  A  E  A  N  G  W  I  E  A

271  ACAAATCCAATGACTCAACTTAAAGGGATGGTCCCCATTAGTTATTTTGAAATATTTGATCGATCTCGTCCTACAGTTACAGCATCATCA
 91   T  N  P  M  T  Q  L  K  G  M  V  P  I  S  Y  F  E  I  F  D  R  S  R  P  T  V  T  A  S  S

361  AACAGTTTTACAAATTCCATTGATATTCAACATCAACATCAACAAGGAATTCACAATGGAACAGGAAATCGAAATTTAAATCAAACATTA
121   N  S  F  T  N  S  I  D  I  Q  H  Q  H  Q  Q  G  I  H  N  G  T  G  N  R  N  L  N  Q  T  L

451  TATGCTGTTACACTATATGAATTTAAAGCTGAACGAGATGATGAATTGGATATAATGCCTAATGAAAATTTAATTATTTGTGCACATCAT
151   Y  A  V  T  L  Y  E  F  K  A  E  R  D  D  E  L  D  I  M  P  N  E  N  L  I  I  C  A  H  H

541  GATTATGAATGGTTTATTGCCAAACCAATAAATCGATTAGGTGGACCAGGTTTAGTACCTGTTTCTTATGTTAAAATAATTGATCTTTTA
181   D  Y  E  W  F  I  A  K  P  I  N  R  L  G  G  P  G  L  V  P  V  S  Y  V  K  I  I  D  L  L

631  AACCCTAATTCTCATTATACATCAATTGATACATCAAGGCGATCACAAGTCATACAAGTAATCAATGGATTTAATATACCGACAGTAGAA
211   N  P  N  S  H  Y  T  S  I  D  T  S  R  R  S  Q  V  I  Q  V  I  N  G  F  N  I  P  T  V  E

721  CAATGGAAAAATCAAACTGCCAAATATCAAGCTTCAACAATCCCCCTTGGTTCAATATCAGGAAGTGGTACTCCACCAACATCAGCTAAT
241   Q  W  K  N  Q  T  A  K  Y  Q  A  S  T  I  P  L  G  S  I  S  G  S  G  T  P  P  T  S  A  N

811  TCACAATATTTTGATAATCATACTATGACTTCAAATCGATCATCACTGGGTTCATCAATTTCTATTATTGAAGCTAGTGTTGATTCATAT
271   S  Q  Y  F  D  N  H  T  M  T  S  N  R  S  S  S  G  S  S  I  S  I  I  E  A  S  V  D  S  Y
```

Fig. 12A

```
01  CAATTAGATCATGGTCGATATCAATATTCAATAACTGCTCGATTAAATAATGGCAGAATAAGATATTTATATCGATATTATCAAGATTTT
01   Q  L  D  H  G  R  Y  Q  Y  S  I  T  A  R  L  N  N  G  R  I  R  Y  L  Y  R  Y  Y  Q  D  F

91  TATGATTTACAAGTGAAATTATTAGAATTATTTCCTTATGAAGCTGGGAGAATTGAAAATTCTAAAAGAATAATTCCATCTATACCAGGA
31   Y  D  L  Q  V  K  L  L  E  L  F  P  Y  E  A  G  R  I  E  N  S  K  R  I  I  P  S  I  P  G

81  CCTTTAATTAATGTCAATGATTCAATATCAAAATTACGAAGAGAAAAATTGGATTATTATTTATCAAATTTAATTGCATTACCTAGTCAT
61   P  L  I  N  V  N  D  S  I  S  K  L  R  R  E  K  L  D  Y  Y  L  S  N  L  I  A  L  P  S  H

71  ATATCTCGATCAGAAGAAGTATTAAAATTATTTGATGTTTTAGATAATGGATTTGATCGAGAAACTGATGCTATTAATAAACGATTTTCT
91   I  S  R  S  E  E  V  L  K  L  F  D  V  L  D  N  G  F  D  R  E  T  D  A  I  N  K  R  F  S

61  AAACCAATAAGTCAAAAATCAAATTCTCATCAAGATAGATTATCTCAATATTCCAATTTTAACGTTTTACAACAACAACAACAACAACAG
21   K  P  I  S  Q  K  S  N  S  H  Q  D  R  L  S  Q  Y  S  N  F  N  V  L  Q  Q  Q  Q  Q  Q  Q

51  CAACAACAGCAATATGCTCATCATTCAAGAGGTTCTGATAATTCACCTACTAATGAATCATCAGGTTCAAATTTAATTAATTCTTCTTCT
51   Q  Q  Q  Q  Y  A  H  H  S  R  G  S  D  N  S  P  T  N  E  S  S  G  S  N  L  I  N  S  S  S

41  CATAATGATTCATCATTATCTTCATCACCACCACCACCACCACCACAAACTGTCACCACCACGAACACcCACGAACACCACCATAACCACA
81   H  N  D  S  S  L  S  S  S  P  P  P  P  P  P  Q  T  V  T  T  T  N  T  T  N  T  T  I  T  T

31  GACTCCTCATCAAAACAACCAAAAGCCAAAGtGAAATTTTATTTTGATGATGATATATTTGTATTATTAATCCCAACCAATTTACGATTA
11   D  S  S  S  K  Q  P  K  A  K  V  K  F  Y  F  D  D  D  I  F  V  L  L  I  P  T  N  L  R  L

21  CAAGATTTAAAATCAAAATTATTTAAACGATTAGAATTGGATATTACTTATAAATATGAAAAACCTGATCAACAACAAAAACCTACATCA
41   Q  D  L  K  S  K  L  F  K  R  L  E  L  D  I  T  Y  K  Y  E  K  P  D  Q  Q  Q  K  P  T  S

11  GAATCAATTCATTTATTTTTGAAAAATGATTTTGAAGATTTTTTAATTGAAAATGAAACTAGCAACAACAACAATCTGGAAATTGATTTC
71   E  S  I  H  L  F  L  K  N  D  F  E  D  F  L  I  E  N  E  T  S  N  N  N  N  S  E  I  D  F

01  GAAAATGAAATTATTAAAGAAAAATTAGGAGAATTTGAAGTTAATGATGATGAAAAaTTTCAAaGTATTTTATTTGATAAATGTAAATTA
01   E  N  E  I  I  K  E  K  L  G  E  F  E  V  N  D  D  E  K  F  Q  S  I  L  F  D  K  C  K  L

91  ATGGTTTTAGTATATTAAACAGAGATCAATAAGAGAGAGAGAGAGAGACAT
31   M  V  L  V  Y  -
```

Fig. 12B

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/CA 97/00809

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/31 C07K14/40 C12N9/12 C12N9/16 C12Q1/18
G01N33/68 //(C12Q1/18,C12R1:725)

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevent passages	Relevent to claim No.
A	WHITE TC ET AL: "Candida albicans secreted aspartyl proteinases: isoenzyme pattern is determined by cell type, and levels are determined by environmental factors." J BACTERIOL, SEP 1995, 177 (18) P5215-21, UNITED STATES, XP002057294 see abstract see page 5125, right-hand column, line 11 - line 15 see page 5220, right-hand column, paragraph 1 - paragraph 2 --- -/--	1,6

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later then the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

4

Date of the actual completion of the international search	Date of mailing of the international search report
2 March 1998	11/03/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Gurdjian, D

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/CA 97/00809

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropnate, of the relevant passages	Relevant to claim No.
A	HUBE B ET AL: "Expression of seven members of the gene family encoding secretory aspartyl proteinases in Candida albicans." MOL MICROBIOL, OCT 1994, 14 (1) P87-99, ENGLAND, XP002057295 see abstract; figures 6-10 ---	1,6
A	GOODVIN, ANDREW R. ET AL: "Purification and characterization of cyclic AMP-dependent protein kinase from Candida albicans" MYCOL. RES. (1996), 100(5), 625-631 CODEN: MYCRER;ISSN: 0953-7562, XP002057296 see abstract see page 625, left-hand column, paragraph 1; figure 5; table 3 ---	1,6
A	NAVARRO-GARCIA, FEDERICO ET AL: "Functional characterization of the MKC1 gene of Candida albicans, which encodes a mitogen-activated protein kinase homolog related to cell integrity" MOL. CELL. BIOL. (1995), 15(4), 2197-206 CODEN: MCEBD4;ISSN: 0270-7306, XP002057297 see abstract see page 2204, left-hand column, line 16 - right-hand column, line 2 see page 2205, right-hand column, last paragraph ---	1,6
A	EP 0 472 286 A (MERCK & CO INC) 26 February 1992 see abstract see page 3, line 33 - line 37 ---	1,6
A	WO 94 28914 A (MITOTIX INC) 22 December 1994 see claims 1-26 ---	1,6
P,X	KOHLER JR ET AL: "CANDIDA-ALBICANS STRAINS HETEROZYGOUS AND HOMOZYGOUS FOR MUTATIONS IN MITOGEN-ACTIVATED PROTEIN-KINASE SIGNALING COMPONENTS HAVE DEFECTS IN HYPHAL DEVELOPMENT" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 1996, 93, 13223-13228, XP002057298 see the whole document ---	6

-/--

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/CA 97/00809

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document. with indication.where appropriate. of the relevant passages	Relevant to claim No.
P,X	LEBERER E ET AL: "SIGNAL-TRANSDUCTION THROUGH HOMOLOGS OF THE STE20P AND STE7P PROTEIN-KINASES CAN TRIGGER HYPHAL FORMATION IN THE PATHOGENIC FUNGUS CANDIDA-ALBICANS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 1996, 93, 13217-13222, XP002057299 see the whole document	6
P,X	WO 97 38129 A (MITOTIX INC ;UNIV JOHNS HOPKINS (US)) 16 October 1997 see page 1 - page 12; claims 23-130; figures SEQ.5,6	1,6
P,X	WO 97 38293 A (MITOTIX INC ;UNIV JOHNS HOPKINS (US)) 16 October 1997 see page 1 - page 12; claims 23-130; figures SEQ.5,6	1,6

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter: nal Application No
PCT/CA 97/00809

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0472286 A	26-02-92	CA 2047028 A	19-01-92
		JP 5068567 A	23-03-93
WO 9428914 A	22-12-94	US 5443962 A	22-08-95
		AU 7055194 A	03-01-95
		EP 0708653 A	01-05-96
		JP 8510650 T	12-11-96
WO 9738129 A	16-10-97	AU 2452097 A	29-10-97
		AU 2726597 A	29-10-97
		WO 9738293 A	16-10-97
WO 9738293 A	16-10-97	AU 2452097 A	29-10-97
		AU 2726597 A	29-10-97
		WO 9738129 A	16-10-97